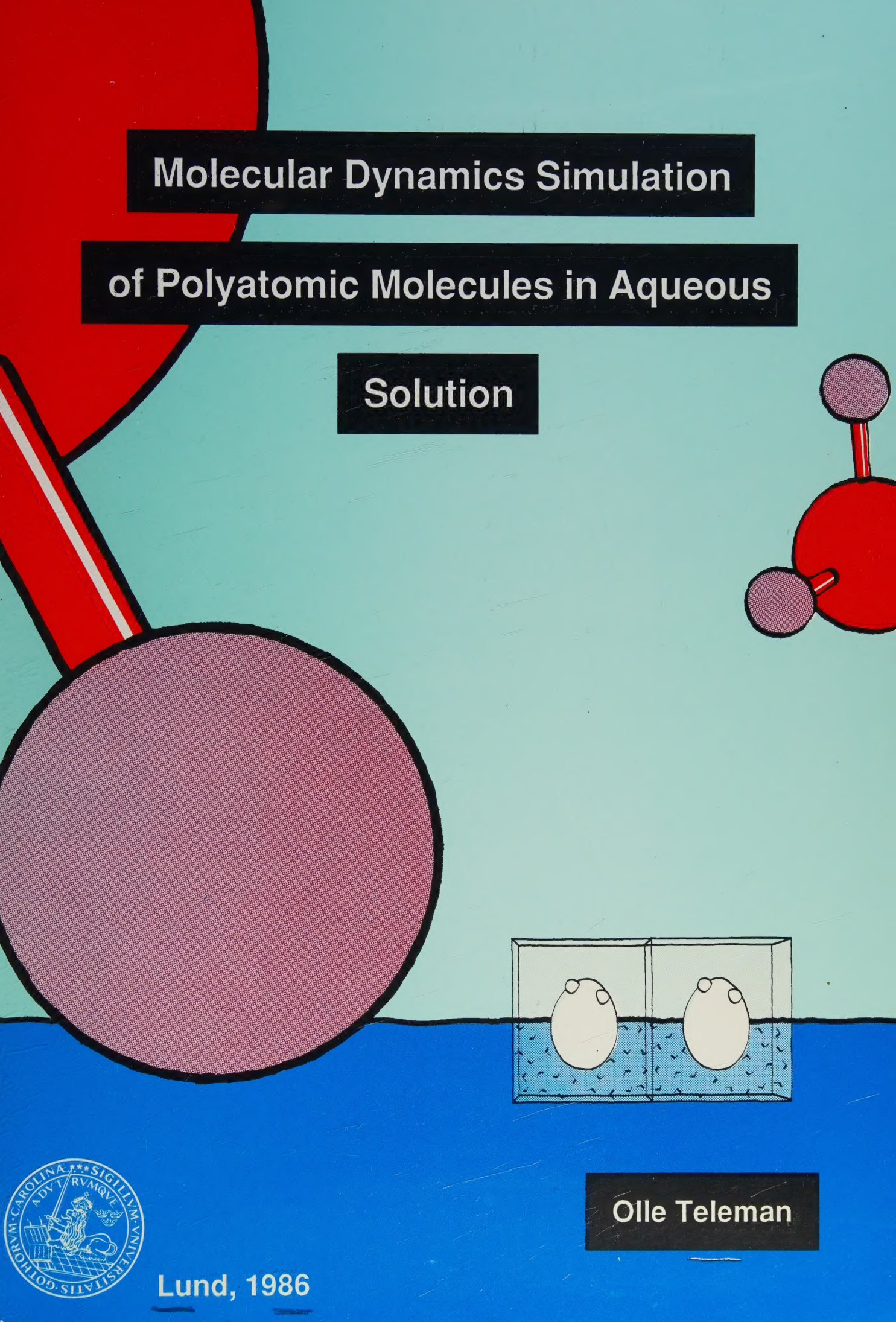


Molecular Dynamics Simulation of Polyatomic Molecules in Aqueous Solution



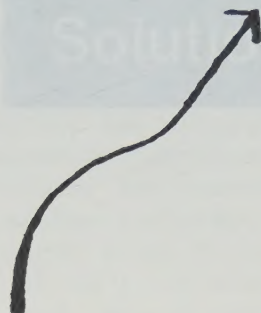
Lund, 1986

Olle Teleman



Molecular Dynamics Simulation

Molecular Dynamics Simulation of Polyatomic Molecules in Aqueous Solution

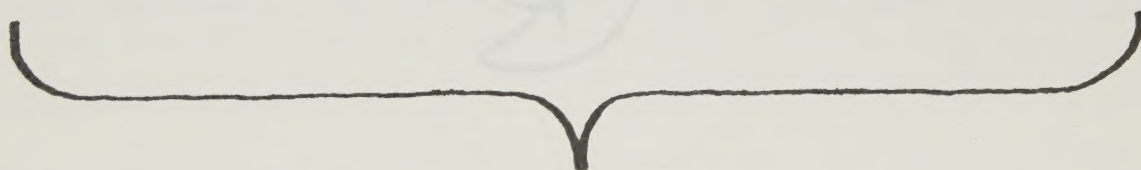


Olle Teleman

Lund, 1986

*Avhandling för teknologie
doktorsexamen.*

*Teknologie licentiat,
Smålands nation.*



*Avhandlingen kommer att försvaras vid en offentlig
disputation, denna äger rum fredagen den 30:e maj 1986, kl 10¹⁵
i sal D på Kemicentrum, Lund.*

Olle Teleman

Lund, 1986



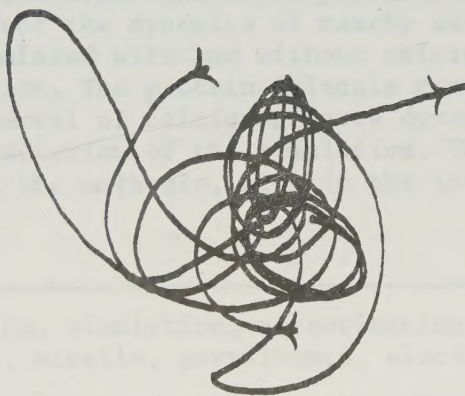
Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41546727_0052

Molecular Dynamics Simulation

of Polyatomic Molecules in Aqueous

Solution



Olle Teleman

Lund, 1986

Molecular Dynamics Simulation

of Polyatomic Molecules in Aqueous

Solution



Bilden på försättsbladet visar två partikelbanor av tre ur en MD-simulering av det klassiska trekropparsproblemet och illustrerar därmed att den numeriska simuleringen ger oss tillträde till områden dit den exakta matematiken ännu inte nått. Bildtexten har medvetet införts på svenska, det är inte mer än rimligt att någon del av avhandlingen är utförd på författarens modersmål.

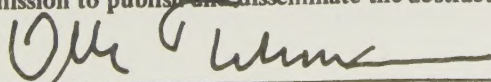
Lund, 1986

Organization LUND UNIVERSITY Dept. of Physical Chemistry 2 Chemical Centre POB 124 221 00 LUND	Document name DOCTORAL DISSERTATION	
Author(s) Olle Teleman	Date of issue May 5th, 1986	
	CODEN:	
Title and subtitle Molecular Dynamics Simulation of Polyatomic Molecules in Aqueous Solution.		
Abstract A general purpose computer program for Molecular Dynamics (MD) simulation of arbitrary mixtures of molecules has been developed. The program is extensively vectorized and uses a double time step algorithm. Execution times compare favourably with literature. Water has been simulated. Intramolecular flexibility accelerates dynamics by a factor of two to three. It is found that intramolecular flexibility can not be treated in an average way. A calcium ion in water has been simulated, and found to have a coordination number of 7. Dynamic properties are consistent with those of the water model. The calcium chelator Ethylene-diammine tetraacetic acid has been simulated with and without calcium, in aqueous solution. The presence of calcium strongly affects thermodynamic and dynamic properties of the chelator. The calcium ion has a coordination number of 7 in the complex. Parallel simulations with different initial configurations show that an equilibration time of 20 ps is adequate. A sodium octanoate micelle in aqueous solution has been simulated. The micelle is maintained without any constraints. Water penetrates into the outer parts of the micelle, which is slightly aspherical. Reorientation of the micelle is an order of magnitude slower than that of the monomers. Reorientational correlation times and order parameters are consistent with NMR data. The micelle does not affect the dynamics of nearby water. The calcium binding protein Parvalbumin has been simulated with and without calcium in vacuo, as well as with calcium in aqueous solution. The protein molecule contracts slightly in the simulations, but less so in water. Removal of calcium affects dynamic properties in all parts of the molecule within the duration of the simulation. The presence of water affects dynamics in all parts of the molecule, also in the interior.		
Key words Molecular Dynamics, simulation, vectorization, double time step algorithm, water, calcium ion, EDTA, micelle, parvalbumin, electrostatics, dynamics.		
Classification system and/or index terms (if any)		
Supplementary bibliographical information		Language English
ISSN and key title		ISBN
Recipient's notes	Number of pages 180 Price	
	Security classification	

Distribution by (name and address) Olle Teleman, Dept. Physical Chemistry,
Chemical Centre, POB 124, S-221 00 Lund

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature



Date April 2nd, 1986

Molecular Dynamics Simulation of Polyatomic Molecules in Aqueous Solution

Summary of Thesis

Introduction

Experimental investigation is our first means of finding out about the physical world around us. The analogy is immediate between the NMR spectroscopist at his (her) sophisticated instrument and the child intently peering at a beetle in the grass. On very large systems, e. g. such as are encountered in astronomy, and very small, e. g. molecular or nuclear, direct experimental evidence may be hard to come by. In such cases, it is sometimes possible to organize available findings into a theory, and thus to widen our knowledge by extrapolation. Occasionally, theory even allows such extrapolations that are very far-reaching.

Theory and experimental knowledge have now reached a state that allows the construction of more or less accurate molecular potential functions, that is, functions that specify the potential energy of a system as a function of the coordinates of the constituent particles. Where experimental evidence is not available, these potential functions can be used to predict static properties. This is achieved by direct energy minimization or by more sophisticated techniques, that take statistics into account. One can also predict dynamic properties, if one makes the additional assumption of an equation of motion, such as that of Newton. The equation of motion normally involves a force, which is obtained as the negative gradient of the potential. The force on each particle depends on the position of very many other particles, and the equation of motion has to be solved numerically, step by step. This procedure is termed Molecular Dynamics (MD) simulation, and the result is a trajectory containing the position of all particles as a function of time. MD simulation is the subject of this thesis, which, apart from this summary, consists of five papers, namely:

- P. Olle Teleman & Bo Jönsson, "Vectorizing a General Purpose Molecular Dynamics Program", *J. Comp. Chem.* **7**, xxx (1986).
- W. Olle Teleman, Bo Jönsson & Sven Engström, "Molecular Dynamics Simulations of a Water Model with Intramolecular Degrees of Freedom".

- E. Olle Teleman & Peter Ahlström, "Molecular Dynamics Simulation of a Small Calcium Complex in Aqueous Solution", J. Am. Chem. Soc., in press.
- M. Bo Jönsson, Olle Edholm & Olle Teleman, "Molecular Dynamics Simulation of a Sodium Octanoate Micelle in Aqueous Solution", submitted.
- Pa. Peter Ahlström, Olle Teleman, Bo Jönsson & Sture Forsén, "Molecular Dynamics Simulation of Parvalbumin in Aqueous Solution".

I will outline the more salient results of these papers, which will be designated by P, W, E, M and Pa. The summary makes use of no references in addition to those of the papers and no separate literature list is provided. References will carry their original numbers preceded by the above-mentioned letter-code. Equations, figures and tables of the individual papers will be referred to similarly.

Potential Model

Molecular Dynamics simulation relies on a potential model, from which a trajectory is generated by means of an equation of motion. The functional form of the potential, as applied here, is given by Eqs. 1 and 2. The potential is based on site-site interactions, where a site is either an atom or a group of atoms (a pseudoatom). Pseudoatoms have been used in the micelle (M) and parvalbumin (Pa) simulations, where the only pseudoatoms are CH, CH₂ and CH₃ groups. Eq. 1 contains the noncovalent contributions, which are assumed to consist of a 6-12 Lennard-Jones potential^{E14} plus an electrostatic term:

$$U_{nc} = \begin{cases} \sum_{i<j} \left[4\epsilon_{ij} \left\{ \frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right\} + \frac{q_i q_j}{4\pi\epsilon\epsilon_0 r_{ij}} \right] & , r_{ij} < r_{cut} \\ 0 & , r_{ij} > r_{cut} \end{cases} \quad (1)$$

where r_{ij} is the distance between two sites i and j . Partial charges q_i and q_j are taken from the

literature^{P6,E15,E16}. The Lennard-Jones parameters ϵ_{ij} and σ_{ij} are calculated from the Kirkwood-Slater formula^{E17} and from the assumption that the Lennard-Jones potential minimum corresponds to the sum of the van der Waals radii of the two sites. Data are taken from refs. E18-E20 and M35. The simple point charge (SPC) water model of Berendsen et al.^{E20} is used. The dielectric permittivity is assumed to be 1. Eq. 1 is taken to describe not only intermolecular interactions, but also those between sites belonging to the same molecule, but separated by at least four covalent bonds (micelle simulations) or three covalent bonds (all other simulations).

The covalent contributions to the potential (Eq. 2) comprise harmonic terms in all covalent bond lengths and angles, as well as periodic terms in all dihedral angles:

$$\begin{aligned}
 U_C = & \sum_{\text{bonds}} A_i (r_i - r_{i,e})^2 + \sum_{\text{angles}} B_j (\alpha_j - \alpha_{j,e})^2 + \\
 & + \sum_{\text{dihedrals}} [C_{k,0} + \sum_{l=1}^3 C_{k,l} \cos(l\beta_k)] \quad (2)
 \end{aligned}$$

where r_i is the actual length of the covalent bond i , $r_{i,e}$ the corresponding equilibrium length, α_j the actual value of the bond angle j , $\alpha_{j,e}$ the corresponding equilibrium angle and β_k the actual value of the dihedral angle k . Parameters B_j , $r_{i,e}$, A_j , $\alpha_{j,e}$ and $C_{k,l}$ are taken from the literature^{P6,E19-22}. $C_{k,0}$ ensures that $U_C \geq 0$.

Simulation Technique

Trajectories are generated from the potential by numerical integration of the Newtonian equation of motion. With the exception of a few simulations of rigid molecules, all systems are treated entirely as particle systems, and all intramolecular degrees of freedom are treated explicitly. The integration is performed by a fourth order predictor and corrector algorithm, where the predictor is a standard Taylor expansion and the corrector uses coefficients according to Gear^{E12}. The oscillation period of the faster vibrational modes amounts to ≈ 10 fs ($1 \text{ fs} = 10^{-15} \text{ s}$) and places a severe upper limit on the integration time step. To avoid any concomitant excessive demand on computer time, a multiple time step technique is used. Two time steps are used, the smaller is 0.2 fs and the larger between 1.0 fs and 2.0 fs.

Crystal structures determined by X-ray measurements are used as starting configurations where appropriate. Water molecules are placed on the bases of a primitive cubic lattice. Rectangular periodic boundary conditions are used except for two parvalbumin simulations, which are performed *in vacuo*. The noncovalent potential is truncated at a certain distance (r_{cut}) in order to reduce the amount of computation. The cutoff distance has to be fairly large to allow for the long ranged electrostatic term in Eq. 1. Thus, a compromise has to be made between accuracy and computational economy. For the smaller systems, system size determines the cutoff distance, while it is set to 9 Å or 10 Å in the larger systems. A neighbour list is used to keep track of interacting sites. Due to the truncation of the noncovalent potential the total energy is no longer a constant of motion. To remedy the drift in energy, velocities are scaled with an interval of typically 100 fs. All systems are equilibrated for fairly long periods of simulation. The time necessary for equilibration is investigated in paper E. The temperature is calculated from the total kinetic energy and the pressure is obtained from the virial, see Eq. E:1.

The simulations of rigid water molecules reported in paper W are performed using the quaternion technique^{W17} and only one time step. The predictor and corrector algorithm of Gear is used as above, but of order 4 for translation and order 3 for rotation.

Simulations are analysed, among other properties, in terms of radial distributions, distance and angle distributions, translation diffusion coefficients and reorientation time correlation functions. Comprehensive outlines of these standard mathematical tools are given in the different papers where they are applied.

Computational Efficiency

Paper P deals with the computer Molecular Dynamics (MD) program, developed expressly for the purpose of simulating arbitrary mixtures of flexible molecules. The MD program is concentrically cyclic and only contains code involved in the solution of the equations of motion. It caters for all details of the simulation technique given in the previous section, except for the simulation of rigid molecules.

The MD program has two features, one of which is its close adaptation to the architecture of a vector computer such as the Cray 1A, on which most simulations reported here have been performed. The program uses different lists extensively, e. g. lists of spatial neighbours, and neighbours separated by one, two or three covalent bonds. This is done in order to minimize search times. The use of lists, however, entails indexed addressing, which prevents the corresponding code from vectorizing. This problem is circumvented by splitting loops into vectorizing and nonvectorizing parts. The former, comprising some 95% of all computation, then performs very well. The latter are made more efficient by judicious use of Cray system subroutines and by

increasing the number of statements in short loops, while simultaneously reducing the number of times the loop is executed. The latter technique is sometimes called "stripmining".

The second feature of the MD program is the use of two different time steps. All force contributions are heuristically classified as slowly varying or rapidly so. All positions and derivatives thereof are divided into two sets of positions etc., whereby the sum of the two positions is the real position. The rapidly varying forces are used for a corrector acting on one set of positions, the slowly varying ones for a corrector acting on the other set. Thus it is possible to evaluate the two force contributions with different time intervals, see Eqs. P:5-8 and Fig. P:1. With bond length and angle forces classified as rapidly varying and all other as slowly varying, this ratio can be 5-10, typically. This is thus an efficient way to handle flexible molecules.

The simulation times compare favourably with those reported in the literature for other general and also specific Molecular Dynamics simulation programs.

Water

I now turn to the simulation of pure water, namely the one discussed in paper E, as well as those reported in paper W.

The first water simulation (trajectory R1 of paper W) is a reproduction of the original SPC water simulation of Berendsen et al.^{E20}, with rigid molecules. Using the same parameters, but with a different algorithm, the results of Berendsen et al.^{E20} are reproduced within statistical fluctuations.

With intramolecular flexibility included properties change (water simulation of paper E, also included into paper W under the designation F). The water molecules are deformed, so that bonds are elongated by 0.016 Å and angles decreased by 4.6° on average. The dipole moment thus increases by ≈7%. Bond length and angle distributions are narrower than for the free molecule. A second simulation of rigid molecules has been performed, but with the average molecular geometry of the F simulation, in order to investigate whether the consequences of intramolecular flexibility can be reproduced in this way (the R2 trajectory of paper W). Now, compared to the radial distributions g_{OO} , g_{OH} and g_{HH} for the R1 and R2 trajectories those for F display considerably less secondary structure. The first maximum and minimum are more pronounced for R2 than for R1, whereas secondary structure is the same or even weaker in the R2 case. The variation of the higher multipole moments with molecular geometry may account for the latter effect.

In the F trajectory, internal degrees of freedom possess potential and kinetic energy. The energy of the bonds is close to the equipartition value, but the energy of the angles is larger than that. The angles thus couple more strongly to intermolecular degrees of freedom. The total potential energy decreases on inclusion of flexibility. The net decrease is of the same order as the estimated

non-additive contribution for the non-empirical water potential due to Matsuoka et al.^{W2,W25}. The pressure increases from 0.7 kbar to 1.7 kbar on inclusion of intramolecular flexibility.

The dipole moment fluctuation density, calculated over a sphere, is about the same for the F and R2 cases, but $\approx 10\%$ larger than for the R1 trajectory. With the Clausius-Mosotti equation, cf. Eq. W:4, this corresponds to a decrease in dielectric permittivity from 80 to 21.

Dynamic properties, but also thermodynamic, are strongly affected by intramolecular flexibility. The translational diffusion coefficient is 40% larger for F than for R1, whereas it is 40% smaller for R2 than for R1. Reorientational correlation times are $\approx 40\%$ shorter for F than for R1, while those for R2 are $\approx 60\%$ longer. The ratio of the correlation time for the first order Legendre polynomial to that of the second is significantly lower than that for a pure diffusion model. The translational diffusion coefficient for the F case is larger than the experimental value by a factor of 2.6, while the rotational correlation time is shorter than the corresponding NMR value^{E33} by an approximate factor of three.

The large differences between the F and R2 simulations point to the severe difficulties met with when trying to treat intramolecular vibrations in an average way.

A Calcium Ion in Aqueous Solution

One calcium ion and 124 water molecules have been simulated, cf. paper E, corresponding to a calcium concentration of 0.45 M. The radial distribution of water molecule oxygen atoms around the calcium ion contains two well resolved neighbour layers. The integral of the first gives a coordination number of 7.0 ± 0.2 at a mean distance of 2.3 Å. The orientation of water molecules liganded to the calcium ion is very strongly polarised, with the orientation density maximum for the water dipole vector parallel to the vector from the calcium ion to the oxygen.

The translational diffusion coefficient is found to be $2.7 \pm 0.4 \cdot 10^{-9} \text{ m}^2/\text{s}$. This is larger than the experimental value by a factor of 2.8, which is very close to the ratio between simulation and experimental values for the water self-diffusion coefficient. The time correlation function for the water ligand dipole vector appears to be determined by the diffusion of the water ligand around the ion, as the correlation function decays about as quickly as that for the vector from the calcium ion to the ligand oxygen. The time correlation functions for the two vectors perpendicular to the dipole vector of the ligands are similar to the corresponding correlation functions for bulk water molecules. Five ligand exchanges occur during the simulation, giving a mean residence life time in the order of 50-60 ps.

Simulations of a Small Calcium Complex

Two simulations of CaEDTA^{2-} (EDTA = Ethylene diammine tetraacetic acid) and one of H_4EDTA , all in aqueous solution, cf. paper E, have been performed. Each system contains 314 or 317 water molecules, corresponding to an EDTA concentration of 0.16 M. The crystal structures for two different EDTA complexes have been used as starting configuration for the simulations (trajectories C and M of paper E). The final configuration in the C trajectory has served as initial configuration for the H_4EDTA simulation (referred to as the M trajectory).

In no case is the calcium ion released. Some changes in ligation occur, such as release of one carboxyl ligand, addition of a water ligand and exchange of a water ligand.

The radial distributions for the carboxyl oxygen atoms, the nitrogen atoms and water oxygen atoms, all with respect to the calcium ion, show pronounced neighbour layers. Oxygen ligands are ≈ 2.3 Å from the calcium ion, whereas the distance between calcium ion and nitrogen atoms is slightly larger, 3.2 Å. The liganded carboxyl groups and water molecules are slightly deformed by their interaction with the calcium ion. The three radial distributions contain all possible candidates for calcium ligation. The coordination number, 6.7, is thus obtained from the integral of their sum.

Deformation of the EDTA molecule is measured by means of R factors, cf. Eq. E:10. The configurations of the C and M trajectories are, on average, not more similar to their own starting configuration than to that of the other. Further, very low R factors are found between C and M configurations, i. e. cases are found where the CaEDTA^{2-} molecule in the C trajectory is very similar to that in the M trajectory. On the other hand, R factors are larger when relating to the C trajectory after the release of one carboxyl ligand. This also applies to R factors relating to the EDTA molecule in absence of calcium (H trajectory). The conclusion is that the simulations very clearly distinguish between the EDTA where the calcium ion is present and that where it is not. The similarities between the M trajectory and the early part of the C trajectory indicate that an equilibration period of 17 ps is adequate.

Reorientational motion is analysed by means of time correlation functions (tcf:s). Caution has to be exercised when calculating correlation times from these tcf:s, as correlation times are of the same order of magnitude as the duration of simulation itself. This numerical problem is dealt with in Eqs. E:11 and E:12. The acetic acid C-H vectors have a reorientational correlation time for the second order Legendre polynomial of ≈ 8 ps in the presence of calcium, and ≈ 4 ps in absence thereof. The correlation time for the entire EDTA complex is ≈ 15 ps with calcium ion and ≈ 6 ps without.

The translational diffusion coefficient of the EDTA complex is obtained from the mean square displacement of its centre of mass and is $\approx 1 \cdot 10^{-9} \text{ m}^2/\text{s}$.

The simulation results in some cases deviate from experiment. This reflects imperfections in

the potential and possibly, but less likely, in the simulation technique. It applies mostly to dynamic properties, but also to some structural. The interaction between the calcium ion and the nitrogen atoms is unexpectedly weak. This is probably due to neglect of electronic polarisability in the potential. The nitrogen atoms each possess a lone pair of electrons, which in physical reality are strongly polarised by the calcium ion. Concomitant to the partial neglect of the inductive term in the calcium-nitrogen relation is an increase in internal mobility. Also the carboxyl groups are polarised by the calcium ion. As the electronic polarisability of the carboxyl group is larger in the direction from one oxygen to the other, the polarisation will increase the energy barrier for rotation of that group. From the energies involved, this affects the rotation probability by several orders of magnitude, and carboxyl group flips, and also carboxyl ligand releases, may be entirely inhibited on the time scale accessible to MD simulation.

That rotational correlation times are shorter for the acetic acid C-H vectors than for the overall tumbling, and the fluctuations in the R factors, indicate considerable internal motion. In consequence, the EDTA complex can not be described by a simple hydrodynamic model.

The motion of the CaEDTA^{2-} complex as a whole is determined by the solvent, i. e. by the dynamic properties of water itself. Comparison of experimental and simulation data on water dynamics is used to correct simulation results on the dynamic properties of EDTA. Such a correction in conjunction with the assumption of restricted internal motion accounts for the discrepancy between the simulation and NMR values for the acetic acid C-H vector correlation times, where the experimental value is ≈ 58 ps.

Simulation of a Sodium Octanoate Micelle

Formation of micelles from amphiphiles is governed by a delicate balance between two competing effects; a free energy term associated with the dissolution of the hydrocarbon tails in water, favouring aggregation, which is counteracted by the head group to head group repulsion. The dielectric properties of water are very important to the latter, at least. To my knowledge, the dielectric permittivity of SPC water has not been determined. The permittivity for a similar water potential due to Matsuoka et al.^{M38} has been reported, though, to be ≈ 35 , roughly half the experimental value for water. There are reasons to believe that the SPC water permittivity is not very much larger than this. From this it is likely that a straightforward application of the potential as given in Eqs. 1 and 2 and Tables M:I and M:II (this simulation will be referred to as the full charge (FC) model) will not lead to proper aggregation. One very simple way to remedy the lack of dielectric screening in the water potential is to scale ionic interactions. To test this approach, a second simulation has been performed with Na^+ and -COO^- charges reduced to half their real values, that is $+0.5$ and -0.5 of an elementary charge. This simulation will be referred to as the reduced charge (RC) model.

The trajectories contain vast amounts of data and the number of possible analyses is correspondingly large. They have been limited to properties that are amenable to experimental determination or illuminate the relation between the RC and FC models or between these models and others reported in the literature.

The radius of the micelle is calculated as the average distance between the centre of mass of the micelle (com) and the head group carboxyl carbons. The micelle is larger and more diffuse in the FC case than for the RC model. This is also seen from individual snap shots from the trajectories, as well as from the total hydrocarbon density as a function of distance from com. In the FC case, the micelle is considerably aspherical. The aspherity is calculated from the principal moments of inertia. The rate of micellar deformation is assessed from the time correlation function of the fluctuations in the moments of inertia. From the decay therein it seems that micelle deformation is governed by monomer diffusion rather than by rotation of the whole aggregate.

Translational diffusion coefficients are 5-10 times faster than the experimental ones^{M14}. As the micelle is neither rigid nor strictly spherical, consideration has to be given how to extract the overall rotation of the micelle. This question is dealt with in Eqs. M:3 and M:4. The overall rotation is found to be an order of magnitude slower than the reorientation of the monomers. The dynamic data are interpreted by means of Stokes formulæ for a sphere in a fluid, and yield effective micellar radii consistent with those obtained above from actual geometry.

The density of different hydrocarbon groups as a function of distance from com does not vary significantly with chain position in the FC case, see Fig. M:4a. In the RC case, there are distinct peaks for the different hydrocarbon distributions, cf. Fig. M:4b. The peak for the C1 (the carboxyl carbon, for numbering see Fig. M:1) distribution is found furthest out and then follow C2, C3, C4, C5 and C6, in that order. There is then a reversal and the maxima of the C7 and C8 distributions are outside the C5 maximum. These results are of certain interest, since they can not be reproduced, even qualitatively, by simple packing models. They also differ from distributions reported in the literature^{M19,M23,M24}, where the micelle was maintained in the absence of water by constraints on head group positions.

The electron density of the system as a function of distance from com is less structured than what is usually assumed in models for the interpretation of scattering data. The increase in electron density at head groups is discernible.

The fraction of trans bonds is $\approx 50\%$ for both potential models. The trans-gauche transition rate is $\approx 0.067 \text{ ps}^{-1}$ in the FC case and $\approx 0.038 \text{ ps}^{-1}$ for the RC model.

The time correlation function for the second order Legendre polynomial for the C-H bond orientation determines the ^{13}C NMR dipole-dipole relaxation. This tcf is normalized to start at unity. Within a picosecond it decays to ≈ 0.6 , due to fast vibrational and librational motion within the same conformational state. Then follows a decay with a time constant in the order of 2 ps. This decay brings the tcf down to 0.05-0.1 within 10 ps, and may be caused by trans-gauche

isomerizations. At longer times the uncertainty in the tcf is larger, but there appears to be further components faster than the rotational diffusion.

NMR measurements of longitudinal relaxation times, T_1 , have been interpreted in terms of an order parameter S as well as a short and a long correlation time^{M47}. Direct comparison of these correlation times to simulation is still beyond reach. Simulation order parameters are ≈ 0.2 for C1 to C6 and ≈ 0.1 for C7 and C8, and they agree fairly well with experiment. Order parameters may also be obtained from the monomer structure, i. e. by relating the C-H bond not to itself but to some reference director. Such order parameters are smaller than those obtained from tcf:s, and also decrease faster with chain position.

Water penetrates into the micelle in the sense that the transition region between the hydrophobic and aqueous phases is rather broad. This is certainly so for the FC model, while less so in the RC case, as seen from the water density as a function of distance from com (see Fig. M:9). Water penetration is also monitored by hydration numbers, here defined as the number of water molecules within a certain distance. Hydration numbers display the same trend for both potential models, with slightly larger hydration numbers for the C1 and C2 atoms compared to the middle of the chain, and with a slight increase in hydration towards the end of the chain.

In order to investigate whether the mobility of water molecules is affected by the presence of the micelle, three spherical concentric shells around the micelle are defined and the motion of water molecules in each of them is analysed separately. The inner radii of the shells are 10, 12 and 14 Å and their thickness is 2 Å. Reorientational correlation times are approximately the same in all three shells. Translation is analysed by means of propagators, i. e. the conditional probability $p(k, t|k_0)$ to find a water molecule in shell k_0 at time t_0 and in shell k at time t_0+t . From the self propagators, where $k=k_0$, it is found that water molecules in the innermost shell diffuse as fast or possibly even faster than in the other shells. The macroscopic diffusion equation can be solved in terms of propagators to give a diffusion coefficient by fitting. Obtained values are consistent with the diffusion coefficient of bulk SPC water.

Sodium ions penetrate into the micelle in about the same way as the water. The average number of sodium ions within 14 Å from com is 10.7 (FC) and 4.8 (RC). With micelle and ion radii corresponding to the RC case and a dielectric permittivity of 80, the HNC approximation gives an average number of ions of 5.8 within 14 Å. Ions close to the micelle are not bound to individual carboxyl groups, as seen from Fig. M:15. The number of sodium ions within 5 Å from each of the carbon atoms is larger in the FC case than in the RC case for all carbons. In both cases, this number decreases monotonically along the chain, which can be understood on the basis of the lower dielectric permittivity of the alkyl chains. The ion ion pair correlation function (RC) has two distinct maxima, one at 3.2 Å and one at 10.7 Å, cf. Fig. M:16.

Simulation of Parvalbumin

Two *in vacuo* simulations of carp muscle parvalbumin have been performed, one with Ca^{2+} ions at the two binding sites and one devoid of calcium. These simulations will be referred to as the VAC and APO simulations, respectively. A third simulation, AQ, has been performed in aqueous solution with 2327 water molecules and 3 sodium ions, the latter to make the system electroneutral. The protein concentration in the AQ simulation is 0.018 M. All simulations have used the X-ray coordinates of ref. Pa55 as initial configuration. The AQ simulation was subject to periodic boundary conditions.

In order to estimate the overall rotational diffusion of the protein in solution (AQ trajectory), the time correlation function for the first order Legendre polynomial is calculated and the corresponding correlation time estimated for several long vectors in the protein. Correlation times between 0.9 ns and 2.5 ns are obtained. The rotational motion of the vector between the two calcium ions corresponds to a correlation time of 1.7 ns. The analysed length of the AQ trajectory is 69.72 ps, which accounts for the fairly large uncertainty in correlation times in the nanosecond region. The translational diffusion coefficient is found to be $2 \cdot 10^{-9} \text{ m}^2/\text{s}$.

The deformation of the protein molecule is monitored by means of the radius of gyration and R factors, but also by direct comparison of distances within the hydrophobic core of the protein. The radius of gyration is 12.8 Å for the parvalbumin in its crystal form. The radius of gyration decreases in all three simulations, but more so in the two *in vacuo* ones. At the end of the APO, VAC and AQ trajectories it is 11.5, 11.6 and 12.0 Å respectively. The smaller radii in the vacuum simulations in part reflect that side chains fold onto the protein surface to maximize favourable van der Waals interactions. Also the backbone itself contracts slightly in all three simulations, as seen from the distances between the amide nitrogens as a function of time.

The backbone in simulation is compared to that in the crystal by means of R factors. They are given in Table Pa:II with a 10 ps interval for all three trajectories, and are on average ≈ 3.2 Å for each of the *in vacuo* ones and 2.5 Å for the AQ simulation.

The distances between all phenyl alanines except Phe 2 and Phe 57 are shown in Table Pa:III and confirm the general picture above. The conclusion is that the properties of the protein molecule are very different for the cases where calcium is present and where it is not, and also very different in the vacuum case from the solution case.

Ramachandran diagrams for the trajectories as well as for the crystal form are presented in Fig. Pa:2. Points are more clustered for the AQ case than for the VAC and APO simulations, and it more resembles the X-ray structure. Density is especially large, naturally enough, in the α -helix region. The standard deviations for the Ψ and Φ distributions follow each other fairly closely in all three simulations, and generally amount to 10-15°. In the AQ trajectory, large standard deviations

are found near the N-terminal and at the end of the B helix. In the VAC simulation, large standard deviations are found in the A helix, between it and the B helix, between the D and E helices and also near the EF binding site. In the APO simulation the backbone is much affected by the removal of calcium ions, especially near the EF binding site.

Backbone dynamics is examined by means of tcf:s for the vectors between subsequent α -carbons. Correlation times are mostly 2-10 times longer in the vacuum trajectories than in the AQ case. The most long-lived component in the tcf corresponds to the overall tumbling of the protein molecule. In the vacuum simulations the corresponding correlation time is infinite, while of the order of a few nanoseconds in solution. The parts of the backbone with short correlation times normally show large standard deviations in the dihedral angles. These parts possibly fluctuate more than other parts, but the reason for the observation can also be a deformation, too slow to be unambiguously inferred from an MD simulation of the present duration.

Hydrogen bonds are not explicitly included into the potential, but rather expected to be maintained by electrostatic and Lennard-Jones interactions. All donor acceptor pairs are examined for hydrogen bonds, the duration of which is determined. Hydrogen bonds are found to be short-lived, with life-times of a few picoseconds, typically.

In crystal parvalbumin, the calcium ion at the CD site has six protein ligands and that in the EF site seven protein ligands and one water ligand. When the calcium ions are removed, structure is lost in the binding sites and mobility increases, as seen in the APO simulation. In the VAC simulation, the crystal water is absent. The calcium ion in the CD and EF sites acquire two and one new ligands respectively, giving a coordination number of ≈ 8 for both sites. The three new ligands are oxygens belonging to carboxyl groups already liganded. The bidental ligandation of carboxyl groups is regarded to be an artefact. Also in water, both sites have a coordination number of ≈ 8 , but here each calcium ion has one water ligand. Exchanges of water ligands occur, which indicates that the binding sites are solvent accessible, as they, indeed, have to be.

Correlation times are estimated for all side chains, except alanines and glycines. A comparison of the two vacuum trajectories shows that removal of calcium ions increases mobility in the entire protein, with a few exceptions. Thus, parvalbumin is capable of fast, global information transfer, which is a prerequisite for the strong calcium binding cooperativity found from experiment^{Pa68}.

Correlation times are shorter in aqueous solution than *in vacuo*. This certainly applies to charged and polar side chains on the protein surface, but also to hydrophobic residues in the protein core. The conclusion is that the presence of water substantially affects the dynamics also of the protein interior. A more detailed analysis of the phenyl alanine rings in the protein core shows that the presence of water accelerates their general dynamics, but even more so their restricted rotation around the C_β - C_γ bond.

Outlook

The next step in the development of MD simulation must be an adaptation to remedy the recognized weaknesses of present potential models. Most potential models have empirical origin, whereas, in a few cases, quantum chemical *ab initio* potentials are available. The empirical and quantum chemical potentials complement each other, and one may hope that, at least for simple systems, they will converge. The molecular polarisability has been neglected heretofore. The nuclear part thereof is now tentatively accounted for by means of intramolecular flexibility, whereas no simple technique to include electronic polarisability has been put forward. Correct treatment of the latter may preclude pair-wise additivity, and the issue requires thorough investigation. A further improvement will be to bring available dynamics data, such as they are, to bear in the construction of potentials. The first choice is, of course, an adjustment of the water model. The refinement of macromolecular potentials will very likely proceed along similar lines, but faces the intricate problem of experimental validation. Once these fundamental difficulties have been overcome, the possible advances in the understanding of molecular mechanisms will be very far-reaching. On this optimistic note I conclude my summary.

Acknowledgement

This thesis carries only my name on its cover. This is ridiculous. Very many people have contributed towards this work, and I will try to express my gratitude. First and foremost comes Anita, who is my wife, my friend and my colleague. Together, we have established a family, led a rich social life, and we have even worked a bit together. One of our mutual projects must not remain unmentioned, I am referring to our ten months old son Johan, but for whose diligence and ingenuity this thesis would have been finished months earlier. My mother and my father have helped in taking Johan off our hands every now and then, as have also Birger, Essie and Lina. My brother Henrik helped me design the cover. He is also a brave man, because he has set out to earn his living as an artist. I would also like to mention my sister Pia, who acted extra-parent for Johan last summer, and who will shortly bear twins.

Going back to my early scientific days, and thinking of time-resolved fluorescence, the names of Per-Olof Eriksson, Lennart Johansson and Göran Lindblom come to mind. Among the NMR people, my gratitude is due to Sture Forsén, Torbjörn Drakenberg, Eva Thulin, Lennart Lindahl, Tomas Andersson, Olle Söderman, Ali Khan, Hans Lilja, Lennart Nilsson and very many more. I also think of Conny Hägg, Jan-Erik Norne and Peter Sellers, with whom I have collaborated in teaching, and of Bodil Forsvik and Eva Hagen, who always assisted me in the way of accounting, tickets, booking, payment and the like.

In my present line of business, I first of all would like mention Bo Jönsson and Peter

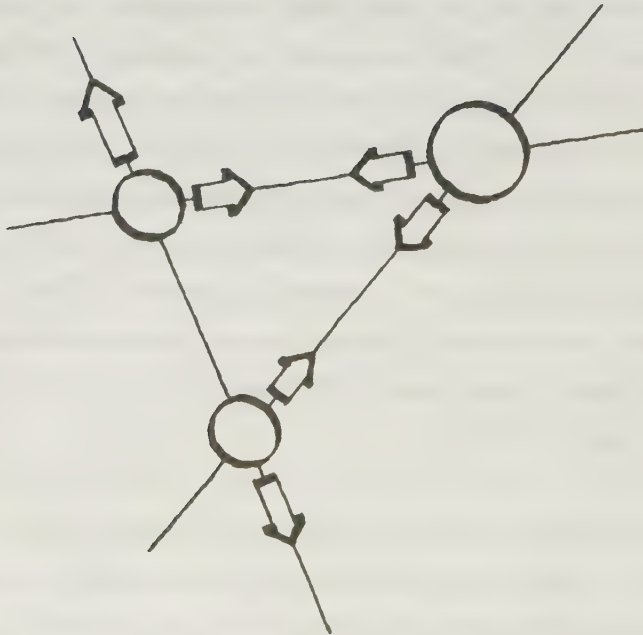
Ahlström. Bo is (a task very nearly over now, possibly) my immediate supervisor and helper, and with him I have worked on all the papers in this thesis. Peter is my colleague and friend, and also the co-author of two of the papers. Immediately connected to the work presented here are also Sven Engström and Olle Edholm. I also think of Marianne Granfeldt, Torbjörn Åkesson, Bo Svensson and Per Linse. Håkan Wennerström was always ready to put his knowledge at my disposal, as was Björn Roos, Bertil Halle, Gunnar Karlström and Bengt Jönsson.

Per-Olof Widmark was always the indispensable computer guru. Roland Lindh provided me with the graphics software, with which some of the figures are drawn. Thanks are also due to the people at Lunds Datacentral and those at Saab-Scania's Datacentral at Linköping, and to the Swedish Natural Science Research Council for generous allocation of computer time.

I always enjoyed the coffee-room lunch community, which was also economic and a sometimes a culinary experience.

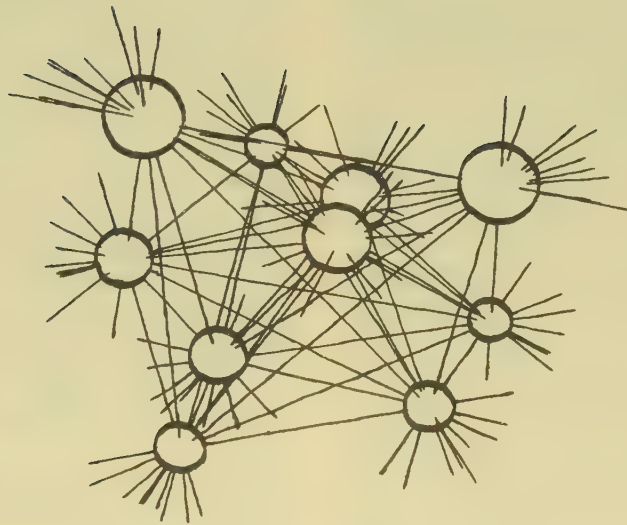
This paragraph, finally, is dedicated to all those whom I have unforgivably forsaken to mention by name.

Lund, March 1986



Pair-wise additivity.

P



Sum over all pairs of i and j .

Vectorizing a General Purpose Molecular Dynamics Simulation Program

Olle Teleman and Bo Jönsson

Dept. of Physical Chemistry 2, Chemical Centre, POB 124, S-221 00 Lund, Sweden

Received 28 January, 1985; accepted 17 June 1985.

A molecular dynamics program for arbitrary molecular mixtures is presented. All intramolecular degrees of freedom are treated explicitly, which means that the program is based on central forces only. A double time step technique has been devised in order to separate rapidly varying, covalent forces from slowly varying ones. Typically, the ratio between the different time steps is about 10, with only a minor computational effort spent in the evaluation of the covalent forces. The program source code is arranged so as to obtain maximal efficiency on a vector processor, while still being portable. On a Cray 1A, a typical simulation of an ion-chelate in aqueous solution with 984 atoms requires a total of 2.9 μ s/interaction with a spherical cutoff distance of 10Å.

INTRODUCTION

Computer simulation techniques play an important role in the statistical mechanical theory of condensed matter.¹ Both the Monte Carlo method of Metropolis et al.² and the solution of Newton's equations of motion—the Molecular Dynamics method—are common numerical techniques today. Since the first Molecular Dynamics (MD) simulation of hard spheres by Alder and Wainwright,³ there has been a steady improvement and refinement of the MD technique. Continuous interatomic potentials of the Lennard-Jones type have been used to study systems such as liquid argon,⁴ and the method has been extended so as to handle rigid polyatomic molecules like water.⁵

To simulate flexible molecules, with certain geometry parameters (most likely bond lengths) kept fixed, is a more difficult problem. The Lagrangian approach with the use of generalized coordinates is straightforward, but fairly tedious and is only applicable to simple molecules.^{6,7} Another technique, which has been widely used, is the so called "method of constraints".⁸ In principle this procedure allows any bond length or bond angle to be held fixed during the simulation. Formally, the constraints enter into the equations of motion as an additional force and require only minor changes in a standard MD program. It is assumed that this is an accu-

rate procedure, as long as the motions of the constrained and unconstrained degrees of freedom occur on well separated time scales. It has been found that constraining bond angles affects the thermodynamics of the system, while using bond length constraints is an acceptable approximation.⁹ However, there are some indications that time-dependent properties can be sensitive even to bond length constraints.¹⁰

In this article we present a different approach based on central forces between all atoms or pseudoatoms (a pseudoatom is a group of atoms treated as a unit). Thus neighbor atoms in the covalent bond structure interact with harmonic forces both in bonds and angles. More distant atoms interact via non-covalent interaction potentials in the usual way.¹¹ With this approach a simulation of a complicated biomolecule in aqueous solution is no more difficult than a simulation of an atomic liquid. A well known drawback of such an approach is that the fast vibrational motions will limit the time step that can be used in the numerical integration. However, this can be overcome by a multiple time step method,^{12,13} where rapidly varying forces are evaluated more frequently than slowly varying ones. In the present program we make use of two different time steps Δt and ΔT , where $\Delta T/\Delta t \approx 5-10$. The advantage of such a division is that the interaction terms consti-

tuting the rapidly varying covalent forces are few and easily evaluated with a small time step Δt . The main computational effort is still spent in the evaluation of noncovalent forces with the large time step ΔT . Thus in a simulation of a biomolecule in aqueous solution the large time step ΔT is not limited by the bond angle or bond length vibrations.

The program to be described in the following sections was originally developed for simulation of biomolecules, but is completely general and applicable to all kinds of mixtures. It is based on a neighbor list technique, and the present version can handle periodic boundary conditions as well as cluster simulations in the microcanonical ensemble. Extensions to other kinds of boundary conditions and ensembles are straightforward.

An MD program for simulation of biomolecules in solution must by necessity lead to large scale computations, which have to be performed on a vector processor. Thus we have taken some care to design the program so that the innermost loops will "vectorize". The vectorization logic is general and not confined to any special machine. As a consequence of the efficient vectorization achieved in the innermost force loop, other parts of the force evaluation become dominant. This has the effect that an increased complexity in the intermolecular potential hardly affects the total cpu time.

The program so far has been used to study pure water, organic chelators, and ionic micelles in aqueous solution, as well as the Ca^{2+} binding protein Parvalbumin. The largest simulation performed comprised 7979 atoms.

We conclude this introductory section by noting that several important issues, such as energy stability and long-range interactions, as well as the choice of numerical algorithm, will be very cursorily treated in this article. These are important aspects of an MD simulation, but the scope of this article has to be limited. Anyway, improved numerical algorithms and special techniques for handling long-range interactions can easily be incorporated in the present vectorized program.

General Outline of the Program

The general structure of the program is cyclic (see Fig. 1) and exclusively contains the

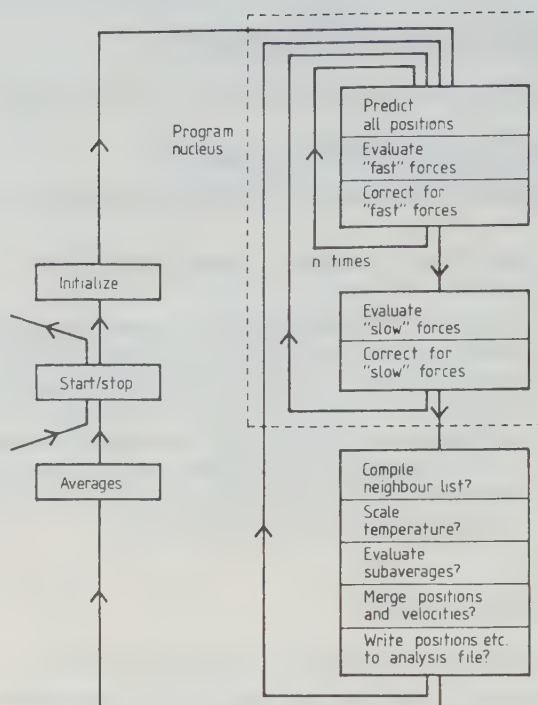


Figure 1. General structure of the MD program. Fast forces are those that vary rapidly. n is the ratio between the two time steps ΔT and Δt . A question mark indicates that the intervals between executions of the corresponding operation can be chosen.

code involved in the actual solution of the equations of motion. All analyses of trajectories, as well as the preparation of the file containing the interaction matrix are done separately in other programs. The program is installed on a scalar ND 500 computer, and on a Cray 1A. The program was developed from an earlier MD program due to Engström¹⁴ written to simulate rigid molecules. The program is available on request from the authors.

To integrate the Newtonian equation of motion the program uses a predictor-corrector algorithm, where the predictor is an ordinary Taylor expansion and the corrector uses coefficients according to Gear.¹⁵ The order of the algorithm may be chosen.

Systems containing any number of pseudo-atom species may be simulated, and these pseudoatoms may be connected with harmonic potentials to form molecules in any way. The total number of pseudoatoms is limited only by the available core memory (see below). The MD program cannot handle rigid molecules, as it relies on central forces. The following two sections will describe the potential used for the simulations, and the subsequent ones will deal with the measures taken to generate fast code.

Covalent Potential

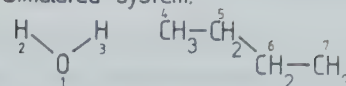
The covalent potential is assumed to be harmonic in bond lengths and bond angles and periodic and even in dihedral angles. Hydrogen bonds and improper dihedral angles⁹ are not explicitly included in the covalent potential, as the former are just Coulombic interactions, and the bond angle potentials take care of the latter. The potential is described by eq. (1), and the parameters are taken from the literature.^{6,9,16-18}

$$V_c = \sum_{\text{bonds}} B_i(r_i - r_{i,e})^2 + \sum_{\text{bond angles}} A_j(\alpha_j - \alpha_{j,e})^2 + \sum_{\text{dihedral angles}} \left(C_{k,o} + \sum_{l=1} C_{k,l} \cos l\epsilon_k \right) \quad (1)$$

Other force fields are available (see ref. 19–28). $C_{k,l}$ is normally 0 for $l > 3$, so that in most cases the summation over l may be truncated after three terms. The constant $C_{k,o}$ ensures that $V_c \geq 0$. Forces are calculated as the gradient of the potential in Cartesian laboratory coordinates. This is trivial for bonds and bond angles, and straightforward but tedious for dihedral angles. An outline is given in ref. 29.

All pseudoatoms in the simulated system are explicitly numbered. From the bond structure of the system, lists are constructed that contain the atoms constituting each bond; similar lists are generated for bond angles and dihedral angles. In accordance with these lists arrays are constructed which explicitly contain the corresponding equilibrium bond distances and force constants, equilibrium bond angles and angular force constants, and the cosine coefficients of the dihedral angle potential, respectively [see eq. (1)]. This is shown diagrammatically in Figure 2 for a simple example. These lists and the corresponding parameter arrays are permanent, as it is assumed that no change in the covalent structure occurs, i.e., no reactions are allowed during the simulations. This extensive listing of covalent potential parameters facilitates the calculation of the corresponding forces, as the program only has to run through these lists. The technique also simplifies the code, but it will still not vectorize entirely, as the program has to access atom positions and store the calculated force contributions by indexed addressing. The

Simulated system:



Potential library:

Bonds	r_e (Å)	B (kJ mol ⁻¹ Å ⁻²)
Water: O - H	100	2318
CH3 - CH2	154	1674
CH2 - CH2	152	1674

Angles:	α (°)	A (kJ mol ⁻¹ rad ⁻²)
Water: H - O - H	109.5	191.5
CH3 - CH2 - CH2	109.5	125.5

Dihedral angle	C0	C1	C2	C3 (kJ/mol)
CH3 - CH2 - CH2 - CH3	209.00	0.0	0.0	209

Resultant explicit arrays:

Bonds	r_e	B
1 2	100	2318
1 3	100	2318
4 5	154	1674
5 6	152	1674
6 7	154	1674

Angles	α	A
2 1 3	109.5	191.5
4 5 6	109.5	125.5
5 6 7	109.5	125.5

Dihedral angle	C0	C1	C2	C3
4 5 6 7	209.00	0.0	0.0	209

Figure 2. This explains how the explicit arrays used to calculate the forces due to the covalent potential are constructed for a system consisting of one water molecule and one butane molecule. Small numbers in the drawing show how the seven pseudoatoms are made up. The system contains five covalent bonds listed in BONDS (1:5, 1:2); therefore, r_e and B each will have five elements containing the corresponding equilibrium bond lengths and potential coefficients. Analogous arrays are used for bond angles and the dihedral angle.

same problem arises for the noncovalent forces, but it can, as we will see below, be partially overcome.

Noncovalent Potential

The noncovalent potential is assumed to consist of a 6-12 Lennard-Jones potential plus an electrostatic contribution, see eq. (2).

$$V_{nc} = \begin{cases} \sum_{\text{pairs}} \left(\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} + \frac{q_i q_j}{4\pi\epsilon\epsilon_0 r_{ij}} \right), & r_{ij} \leq r_{\text{cut}} \\ 0, & r_{ij} > r_{\text{cut}} \end{cases} \quad (2)$$

Partial charges q_i are taken from the literature (refs. 6, 30, 31, and others). The Lennard-Jones coefficients A_{ij} and B_{ij} are calculated from the Kirkwood-Slater formula,³² with parameters from refs. 9 and 32. A large variety of water models is available (see ref. 16 and references therein). The Simple Point Charge (SPC) model of Berendsen et al.¹⁶ is parametrized according to eq. (2) and therefore has been used. Noncovalent potential parameters for many different pseudoatom species are also available in refs. 33–36.

V_{nc} [Eq. (2)] is used to describe not only the intermolecular interactions, but also the interactions between atoms belonging to the same molecule, but separated by a small number of covalent bonds, normally three or four. The relative dielectric constant ϵ is assumed to be 1, since all partial charges are treated explicitly.

The truncation of V_{nc} at $r_{ij} = r_{cut}$ greatly reduces the numbers of pairs to be evaluated, but necessitates keeping track of those pairs ij for which $r_{ij} < r_{cut}$. The MD program uses a neighbor list to do this.³⁷ Setting r_{cut} to 10 Å in an aqueous environment means that interactions with some 135 neighbor water molecules, or about 400 atoms, must be considered. For a system of 1000 atoms, one thus has to list 200 000 neighbors, if reciprocity is used. The neighbor list does not have to be compiled every time step, but the use of a large r_{cut} is a prerequisite for a long interval between compilations. For a cutoff distance of 10 Å and an assumed overlarge value of $8 \cdot 10^{-9}$ m²/s for the water diffusion constant (at 300 K), one finds that about three water molecules enter or leave the neighbor volume in 5 fs. (We have set the interval between neighbor list compilations to 4.8 fs.) This is only a very small part of the total of about 135 water molecules in the neighbor volume. The value of $8 \cdot 10^{-9}$ m²/s for the diffusion constant of water is much larger than the experimental value, but simulations of SPC water yield diffusion constants too large.³⁸

The truncation of the noncovalent potential at $r = r_{cut}$ causes a drift in temperature. To remedy this, the temperature is evaluated at intervals and all velocities are uniformly scaled to attain the preset temperature. Use of a switching function⁴² to truncate the potential smoothly would eliminate the temperature drift. Vectorization properties would not change, but computing time would increase by some 10–30%, depending on the order of the algorithm. While appealing, this technique is not universally of great advantage.⁴³

In order to avoid ionic effects at the cutoff distance, water molecules are considered to be either entirely inside or entirely outside the neighbor volume. In consequence, water hydrogens may be excluded from the neighbor list (see below).

Nearly all of the computing time is spent in calculating the force contributions from eq. (2). To do this for one pseudoatom pair involves the calculation of r_{ij}^{-2} , r_{ij}^{-6} , r_{ij}^{-12} , r_{ij}^{-1} (which implies a square root), $q_i q_j$, and more. For optimal speed these floating point operations have to be done in fully vectorized loop, i.e., it cannot be done directly by using a column of the neighbor list as an index vector. Instead, the calculation is divided into three parts: (i) Collect neighbor positions and parameters into temporary contiguous arrays, which is a (nonvectorizing) “gather” operation; (ii) loop over these arrays to calculate a temporary contiguous force contribution array (vectorized operations); and (iii) add forces from (ii) to the correct pseudoatom total forces, which is a (nonvectorizing) “scatter” operation. The following section will deal with Parts (i) and (iii), and the subsequent one with Part (ii).

Gathering and Scattering

Part (i) uses the column i of the neighbor list to get the numbers j designating the neighbors of atom i , their positions, charges, and Lennard-Jones parameters A_{ij} and B_{ij} . Further, whenever a water oxygen is found in the neighbor list, two corresponding hydrogens are to be included as well. Straightforward Fortran code to do this will be slow, but can be improved upon by the following methods:

(a) Use of the ANINT (Fortran) or CVMGx (Cray Vector Merge)³⁹ functions will allow the code that provides for the periodic boundary conditions to vectorize. In consequence it is put in a loop of its own:

```
DO 14 J = 1,N
  FLX(J) = -XEDGE2 * ANINT(XEDG2I *
    & [XI-XV(J)])
  FLY(J) = -YEDGE2 * ANINT(YEDG2I *
    & [YI-YV(J)])
  FLZ(J) = -ZEDGE2 * ANINT(ZEDG2I *
    & [ZI-ZV(J)])
14 CONTINUE
```

where XI is the x coordinate of atom i , XV(J) those of its neighbors, XEDGE2 the x side of the box, XEDG2I its inverse, and N the number of nonwater neighbors j . The calculated vectors FLX(J) are added to XV(J) before cal-


```

DO 3 J=1,NBMAX
  XDJ=XI-XV(J)+FLX(J)
  YDJ=YI-YV(J)+FLY(J)
  ZDJ=ZI-ZV(J)+FLZ(J)
  R2J=1./((XDJ**2+YDJ**2+ZDJ**2))
  RJ=SQRT(R2J)
  R6J=R2J**R2J
  UH(J)=QIF*VQ(J)*RJ+(C6(J)+C12(J)*R6J)*R6J
  FHJELP=(QIF*VQ(J)*RJ+(6*C6(J)+12*C12(J)*R6J)
    *R6J)*R2J
  VFX(J)=FHJELP*XDJ
  VFY(J)=FHJELP*YDJ
  VFZ(J)=FHJELP*ZDJ
3 CONTINUE
  UHELP=SSUM(NBMAX,UH,1)
  FX1=SSUM(NBMAX,VFX,1)
  FY1=SSUM(NBMAX,VFY,1)
  FZ1=SSUM(NBMAX,VFZ,1)
  ...

```

Figure 3. A well vectorized loop, for the calculation of noncovalent forces. Atom i has NBMAX neighbors j . XI is the x coordinate of atom i , XV(J) those of its neighbors. FLX(J) takes care of periodic boundary conditions in x . VFX(J) is the negative force on neighbor j in the x direction (analogously for y and z). The potential energy is obtained by summing UH(J) and the force on atom i by summing VFX, VFY, and VFZ. This is done with the Cray system subroutine SSUM.

culating the distance between i and j (see Fig. 3). The y and z coordinates are dealt with in an analogous way.

(b) If the subroutine that generates the neighbor list puts water neighbors at the end of each column and returns a variable to specify where they begin, it is possible to loop over nonwater and water neighbors separately; this enables water hydrogens to be included without testing. The increment of the loop in (a) will now be 3 and the loop itself will contain six further statements of the type $\text{FLX}(J + 1) = \text{FLX}(J)$

as the periodic boundary conditions are to apply in the same way to all three atoms in the water molecule. These statements amount to an apparent vector dependency,³⁹ but as this dependency is not real, vectorization can be forced by compiler directives.

(c) Nonvectorized loops have severe loop overhead times, which are comparable to the execution time required for two to four Fortran assignment statements. Thus loops containing only a few statements should be avoided. "Stripmining" of nonvectorized loops will also help, as shown by the following example.

```

REAL A(1000),B(1000)
INTEGER IND(1000)

```

```

C
DO 1 I = 1,1000
  A(IND)(I) = B(I)
1 CONTINUE

```

```

DO 1 I = 1,1000,2
  A(IND)(I) = B(I)
  A(IND)(I + 1) = B(I + 1)
1 CONTINUE

```

Replacement of (4) by (4') will cut Cray 1A execution time by 38%. Stripmining using an increment of 3 will save 51%.

Parts (i) and (iii) may also be written very compactly with extensive use of the Cray Fortran (CFT) system subroutines GATHER and SCATTER, but this code is slower than the one already outlined. GATHER and SCATTER are about 3.5 times faster than the explicit one statement loop Fortran code for an array of 1000 elements, but this is more than overcome by gathering more than one array in the same loop and by stripmining.

A Well-vectorized Loop

Part (ii) consists of one loop over contiguous arrays. This loop vectorizes well³⁹ (see Fig. 3). On a Cray 1A this code has been timed at 50 ± 2 Mflops/s, where an estimate of four floating point operations have been used for the square root. The simulated system contained 984 atoms under periodic boundary conditions and at liquid density; r_{cut} was set to 10 Å. This speed does not vary much with the composition of the simulated system, but increases when r_{cut} increases and vice versa.

Core Memory Limitations

The available core memory determines the maximal size of the systems that can be simulated. With 200 listed neighbors for each atom the program needs about 280 words of core memory per atom. This figure depends slightly on the covalent bond structure of the system, since more memory per atom is required if the bond structure is complex.

If the system consists mainly of solvent molecules, only one atom in each neighboring solvent molecule needs to be included in the neighbor list, i.e., for an aqueous system the neighbor list is reduced by eight-ninths. The core memory requirement is then some

120 words per atom. On a computer with 1 Mword of core, the maximal system size is 3000–8000 atoms.

Memory demand can be reduced by coding the neighbor and covalent lists, or by using half word or bit techniques. However, the concomitant decoding will slow down the program, and a conflict arises between speed and memory economy. Program speed was considered to be more important for our simulations and consequently we have not packed the lists.

Use of Multiple Time Steps

The maximal allowable time step is regulated by the assumption of constant forces, i.e., by the decay of the force autocorrelation functions. In most cases this depends on bond forces in a bond involving a hydrogen atom. The oscillation period of these vibrations is of the order of 10^{-14} s, and a time step of <0.2 fs is necessary to integrate the vibrational motion accurately. The noncovalent forces vary much more slowly and a multiple time step method allows the use of a longer interval between their computation. The evaluation of noncovalent forces is very time-consuming hence the multiple time step method reduces the computer time requirement.

The force on a particle can arbitrarily be divided into two components, and from these two components we may write

$$\mathbf{F}_i(\mathbf{R}) = \mathbf{F}_{i,a}(\mathbf{R}) + \mathbf{F}_{i,b}(\mathbf{R})$$

$$\mathbf{a}_{i,a} \triangleq \frac{1}{m_i} \mathbf{F}_{i,a}(\mathbf{R}), \mathbf{a}_{i,b} \triangleq \frac{1}{m_i} \mathbf{F}_{i,b}(\mathbf{R}) \quad (5)$$

where i refers to any atom in the system, $\mathbf{r}_i$ is the position of that atom and $\mathbf{R} = \mathbf{r}_1, \dots, \mathbf{r}_N$, where N is the number of atoms in the system. It follows that

$$\frac{1}{m_i} \mathbf{F}_i(\mathbf{R}) = \frac{1}{m_i} [\mathbf{F}_{i,a}(\mathbf{R}) + \mathbf{F}_{i,b}(\mathbf{R})]$$

$$= \mathbf{a}_{i,a} + \mathbf{a}_{i,b} = \mathbf{a}_i = \ddot{\mathbf{r}}_i \quad (6)$$

Thus the acceleration $\mathbf{a}_i$ equals the sum of $\mathbf{a}_{i,a}$ and $\mathbf{a}_{i,b}$. It is also clear that $\mathbf{r}_i$ can be divided in such a way that

$$\mathbf{a}_{i,a} = \ddot{\mathbf{r}}_{i,a}, \mathbf{a}_{i,b} = \ddot{\mathbf{r}}_{i,b}$$

$$\text{with } \mathbf{r}_{i,a} + \mathbf{r}_{i,b} = \mathbf{r}_i \quad (7)$$

This is not unambiguous! Any contribution to $\mathbf{r}_i$ that is linear in time can be assigned either

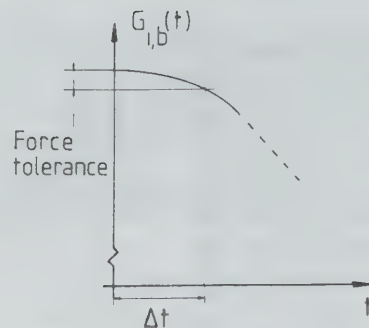


Figure 4. The autocorrelation function $G_{i,b}(t)$ of the most rapidly varying force is chosen to determine Δt . The larger the force tolerance, the less accurate the simulated trajectories will be.

to $\mathbf{r}_{i,a}$ or $\mathbf{r}_{i,b}$ without violating eq. 7. We have

$$\ddot{\mathbf{r}}_{i,a} = \frac{1}{m_i} \mathbf{F}_{i,a}(\mathbf{R}), \ddot{\mathbf{r}}_{i,b} = \frac{1}{m_i} \mathbf{F}_{i,b}(\mathbf{R}) \quad (8)$$

which amounts to a system of $2N$ coupled differential equations, instead of the usual system of N equations.

The Taylor expansion has the property that taking n steps of Δt is equal to taking one step of $n \cdot \Delta t$. Furthermore, if Δt is short compared with the variations in force, the corrections applied by the Gear algorithm are very small. Using this and assuming $\mathbf{F}_{i,b}$ to vary faster than $\mathbf{F}_{i,a}$ a scheme comprising the two innermost loops of Figure 1 is obtained. Now Δt is obtained from $G_{i,b}(t) = \langle \mathbf{F}_{i,b}(t) \cdot \mathbf{F}_{i,b}(0) \rangle$, as in Figure 4, where the most rapidly decaying $G_{i,b}(t)$ should be selected. Application of the same technique to the most rapidly decaying $G_{i,a}(t) = \langle \mathbf{F}_{i,a}(0) \rangle$ yields an upper limit of ΔT , $\Delta T = n \cdot \Delta t$. In practice, n may have to be chosen smaller than that. The division of $\mathbf{F}_i$ into $\mathbf{F}_{i,a}$ and $\mathbf{F}_{i,b}$ is heuristic and as a consequence in general there will be a coupling between the two force components.

This approach differs from that of Swindoll and Haile⁴⁴ in that they expand the slowly varying force component itself in time and not a corresponding set of positions.

We have investigated two force subdivisions according to eq. (5). One of these subdivisions considers the covalent bond forces (due to the first term on the right-hand-side of eq. (1)) to vary rapidly. $\mathbf{F}_{i,a}$ is, therefore, made up of all the other terms in eqs. (1) and (2). Force autocorrelation functions for this force subdivision are plotted in Figure 5(a) for water oxygens. The oscillation in $G_{i,b}(t)$ corresponds to the bond vibration, that in $G_{i,a}(t)$ to

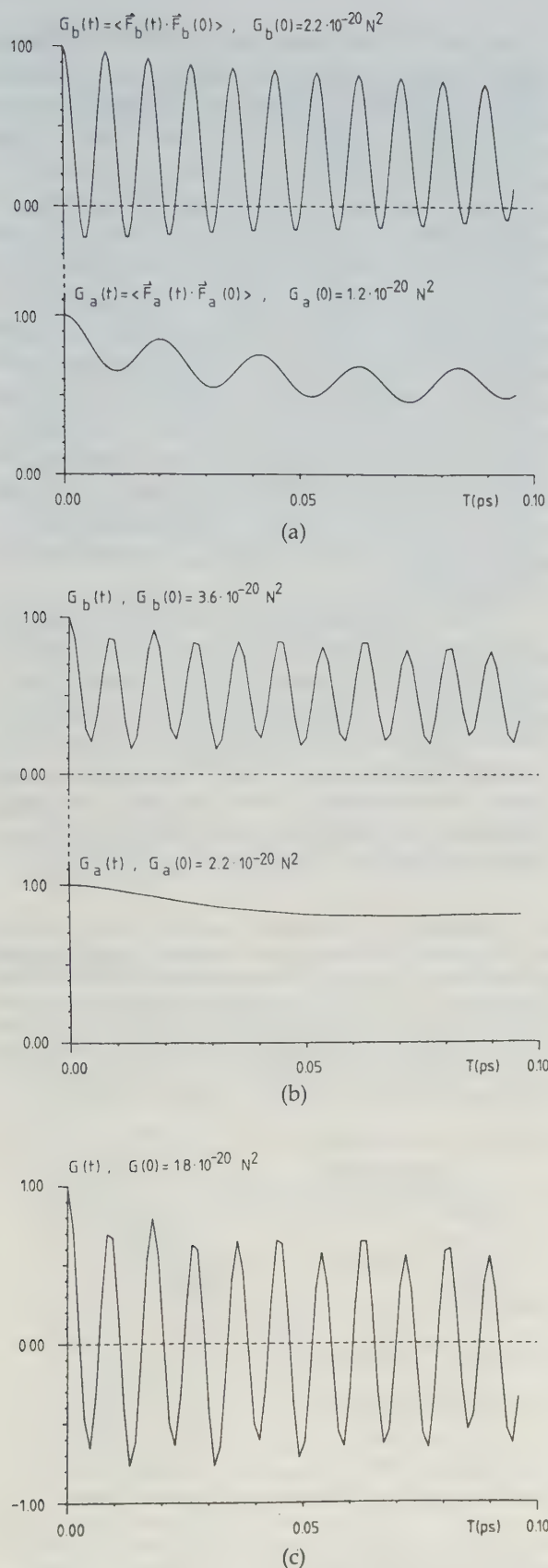


Figure 5. Force autocorrelation functions for bulk water oxygen atoms. (a) For covalent bond forces [$G_b(t)$] and all other force contributions [$G_a(t)$]; (b) for covalent bond and bond angle force contributions [$G_b(t)$] and all other force contributions [$G_a(t)$], i.e., non-covalent forces; and (c) for the total force.

the bond angle vibration. The latter has a period of only slightly more than twice that of the former, and a value of $n > 5$ is clearly inadequate. This is corroborated by the fact that the drift in total energy increases as soon as n is set larger than 5, when $\Delta t = 0.2$ fs. The other force subdivision assigns both bond and bond angle forces (from the first and second terms on the right-hand-side of eq. (1)) to the fast component $\vec{F}_{i,b}$. The corresponding $G_{i,b}(t)$ and $G_{i,a}(t)$ are shown in Figure 5(b), and it is obvious that $n = 8$ or even larger will be satisfactory. The total force autocorrelation function [see Figure 5(c)] is dominated by the covalent forces; this could also have been inferred from the values for $\langle |\vec{F}_i(0)|^2 \rangle$ in Figure 5(b). The maximal values for Δt and n also depend on the composition of the simulated system. If the latter is devoid of explicit light atoms, i.e., hydrogens, Δt can be chosen to be larger. This applies to some *in vacuo* simulations of macromolecules and lipid simulations.

The multiple time step method still can be easily applied if the force is separated into more than two components.

Owing to the ambiguity in the separation, eq. 7, the two solutions $\vec{R}_a$ and $\vec{R}_b$, as well as $\dot{\vec{R}}_a$ and $\dot{\vec{R}}_b$, will wander away in opposite directions. To prevent eventual numerical problems, $\vec{R}_b$ is transferred to $\vec{R}_a$ and $\dot{\vec{R}}_b$ to $\dot{\vec{R}}_a$ at long intervals. This phenomenon is small but noticeable on a 32 bit computer, but almost nonexistent on the Cray 1A, which uses 64 bit words.

The question arises whether a separation of short-range and long-range contributions to the noncovalent forces will allow an even larger time step for distant forces. Correcting less frequently for distant forces enables the use of a larger cutoff distance. In principle this is a feasible strategy. Examples of autocorrelation functions of long-range and short-range forces are plotted in Figure 6. Application of the procedure of Figure 4 would seemingly indicate that the time step for evaluation of distant electrostatically dominated forces only may be chosen slightly larger than that of the short-range forces. As the contribution to the total force is much smaller from distant forces than from short-range forces, a larger time step will not cause a corresponding inaccuracy in the calculated forces. This argument is based on one case

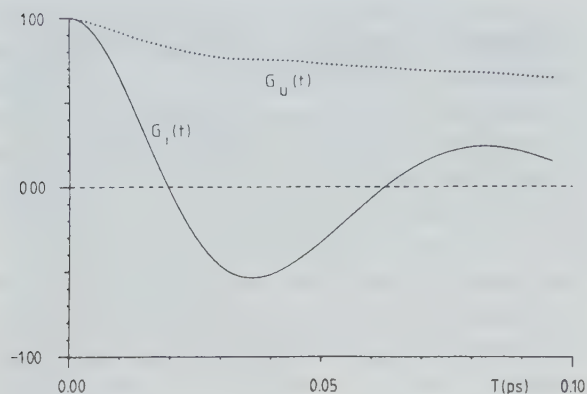


Figure 6. Autocorrelation functions for proximal and distant forces on a calcium ion in a simulation of Ca^{2+} + Ethylene-diammine tetraacetic acid (EDTA) and 317 water molecules. $G_i(t)$ corresponds to forces due to neighbors closer than 8 Å and $G_u(t)$ to those due to neighbors further away than 8 Å but closer than 10 Å; $r_{\text{cut}} = 10$ Å.

only, however, and a generalization may not be straightforward.

Simulation Times

Even with the use of multiple time step techniques, most of the computing time is spent evaluating noncovalent forces. The tim-

ings of three sample simulations are presented in Table I, which also includes force evaluation times per interaction pair. This figure compares favorably with those previously reported in the literature.^{37,40,41} From Table I it is clear that the program is preeminently suited for simulation of aqueous solutions under periodic boundary conditions. For systems such as Parvalbumin *in vacuo* the number of noncovalent neighbor pairs is relatively small and the program will spend comparatively more time in its less well optimized parts. In the Parvalbumin simulation only 32% of the cpu time is spent evaluating noncovalent forces and, therefore, more attention should be given to the bond angle force calculation and the neighbor list compilation.

While it is important that the forces are evaluated efficiently, the choice of algorithm is of equal importance. We are aware that, in some cases, other algorithms may allow the use of larger time steps than those of the present simulations. However, the Gear algorithm is exceedingly well suited to multiple time step algorithms, and thus for the accurate treatment of disparate time scales.

Table I. Cray 1A execution times for three sample systems. Water refers to 216 water molecules under periodic boundary conditions and at experimental density. The system contained 648 atoms, 432 covalent bonds, 216 bond angles and no dihedral angles; $r_{\text{cut}} = 8.5$ Å. Parvalbumin was simulated *in vacuo*. The system contained 983 atoms, 993 covalent bonds, 1429 bond angles, and 1871 dihedral angles; $r_{\text{cut}} = 10$ Å. The micellar system contained a micelle comprising 15 sodium octanoate molecules surrounded by 1094 water molecules. Periodic boundary conditions were applied. The system contained 3447 atoms, 2323 covalent bonds, 1229 bond angles, and 105 dihedral angles; $r_{\text{cut}} = 10$ Å. In all three simulations $F_{i,b}$ consisted of covalent bond and bond angle forces. Numbers in parentheses are percent of total cpu time. The cpu time per interaction is given in microseconds for the different forces.

	Water	Parvalbumin	Micelle
Number of neighbor pairs	82900	83860	619670
Neighbor list interval	24 (4.8 fs)	24 (4.8 fs)	24 (4.8 fs)
No. of time steps (Δt)	48000 (9.6 ps)	1920 (0.384 ps)	960 (0.192 ps)
Total cpu time	2125.22 s	223.00 s	297.55 s
$n = \Delta T / \Delta t$	6	6	6
Predict	108. s (5.1)	6.6 s (3.0)	11.4 s (3.8)
Force evaluation	1705. s (80.3)	136.6 s (61.3)	244.8 s (82.3)
Correct	141. s (6.7)	7.2 s (3.2)	17.0 s (5.7)
Neighbor list compilation	153. s (7.2)	72.0 s (32.3)	24.3 s (8.2)
Temperature scaling	0.2 s (0.0)	0.0 s (0.0)	0.0 s (0.0)
Write, merge	15. s (0.7)	0.5 s (0.2)	0.0 s (0.0)
Forces: Noncovalent	1423. s (67.1)/ 2.2 μ s	71.5 s (32.1)/ 2.7 μ s	212.4 s (71.5)/ 2.1 μ s
Bonds	92. s (4.3)/ 4.4 μ s	8.5 s (3.8)/ 4.5 μ s	10.0 s (3.4)/ 4.5 μ s
Bond angles	187. s (8.8)/18.1 μ s	45.8 s (20.6)/16.7 μ s	21.7 s (7.3)/18.4 μ s
Dihedral angles	0. s (0.0)/ —	10.5 s (4.7)/18.5 μ s	0.4 s (0.1)/25.6 μ s
Noncovalent (i) "gather"	478. s (22.8)	31.9 s (14.5)	73.8 s (25.0)
forces: (ii) "calculate"	586. s (28.0)	24.1 s (11.0)	85.2 s (28.9)
(iii) "scatter"	342. s (16.3)	14.6 s (6.6)	51.8 s (17.6)
Total time/neighbor pair ev.	3.2 μ s	8.3 μ s	3.0 μ s
Ps/cpu hour	16.2	6.20	2.32

CONCLUSION

The care taken to adapt the source code of the MD program has resulted in absolute code that is very well vectorized and, therefore, very fast. The reported execution times are even shorter than those reported for corresponding routines written in assembler code,³⁷ and this emphasizes the inherent potential of the Fortran language, at least for the present needs. The other main feature of this MD program is the use of two different time steps, which, in exchange for a marginally increased core memory requirement, facilitates the investigation of phenomena that occur simultaneously but on different time scales. The multiple time step technique allows expeditious simulation, while also including an explicit description of all intramolecular degrees of freedom.

Note added in proof: Further program refinement and a new Fortran compiler have reduced computing times by about 25% compared to the figures given in Table I.

We want to thank Dr. Sven Engström for placing the SEMD program at our disposal and for his advice and friendly help.

References

1. J.-P. Hansen and I. R. MacDonald, *Theory of Simple Liquids*, Academic Press, London, 1976.
2. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).
3. B. J. Alder and T. E. Wainwright, *J. Chem. Phys.*, **27**, 1208 (1957).
4. A. Rahman, *Phys. Rev.* **136**, A405 (1964).
5. A. Rahman and F. H. Stillinger, *J. Chem. Phys.*, **55**, 3336 (1971).
6. J.-P. Ryckaert and A. Bellemans, *Chem. Phys. Lett.*, **30**, 123 (1975).
7. J.-P. Ryckaert and A. Bellemans, *Faraday Discuss. Chem. Soc.*, **66**, 95 (1978).
8. J.-P. Ryckaert, G. Cicotti, and H. J. C. Berendsen, *J. Comp. Phys.*, **23**, 327 (1977).
9. W. van Gunsteren and M. Karplus, *Macromolecules*, **15**, 1528 (1982).
10. See Figures 5 and 8 of ref. 9.
11. B. R. Brooks, R. E. Brucoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comp. Chem.*, **4**, 187 (1983).
12. W. B. Street, D. J. Tildesley, and G. Saville, in *Computer Modeling of Matter*, P. Lykos, Ed., ACS Symposium Series 86, Washington, 1978.
13. W. B. Street, D. J. Tildesley, and G. Saville, *Mol. Phys.*, **35**, 539 (1978).
14. Program available from S. Engström, Dept. Physical Chemistry, Univ. of Umeå, S-901 87 Umeå, Sweden.
15. C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice Hall, Englewood Cliffs, N.J., 1971, p. 148–154.
16. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces*, B. Pullman, Ed., D. Reidel, Dordrecht, 1981, p. 331–342.
17. D. Dolphin and A. E. Wick, *Tabulation of Infrared Spectral Data*, Wiley Interscience, New York, 1977.
18. G. Herzberg, *Molecular Spectra and Molecular Structure: Infrared and Raman Spectra of Polyatomic Molecules*, Van Nostrand, Princeton, 1945.
19. M. Levitt, in *Protein Folding*, R. Jaenicke, Ed., Elsevier/North Holland, Amsterdam, 17, 1980.
20. M. Levitt, *J. Mol. Biol.*, **82**, 393 (1974).
21. A. Warshel, M. Levitt, and S. Lifson, *J. Mol. Spectrosc.*, **33**, 84 (1970).
22. S. Karplus and S. Lifson, *Biopolymers*, **10**, 1973 (1971).
23. P. J. Rossky, M. Karplus, and A. Rahman, *Biopolymers*, **18**, 825 (1979).
24. J. M. Blaney, P. K. Weiner, A. Dearing, P. A. Kollman, E. C. Jorgensen, S. J. Oatley, J. M. Burrig, and C. C. F. Blake, *J. Am. Chem. Soc.*, **104**, 6244 (1982).
25. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765 (1984).
26. F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).
27. S. Lifson and A. Warshel, *J. Chem. Phys.*, **49**, 5116 (1968).
28. A. T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.*, **96**, 5319 (1974).
29. P. H. Berens and K. R. Wilson, *J. Comp. Chem.*, **4**, 313 (1983).
30. J. A. McCammon, P. G. Wolynes, and M. Karplus, *Biochemistry*, **18**, 927 (1979).
31. J. Hermans, H. J. C. Berendsen, W. F. van Gunsteren, and J. M. Postma, *Biopolymers*, **23**, 1513 (1984).
32. H. Margenau and N. R. Kestner, *Theory of Intermolecular Forces*, Pergamon Press, New York, 1969.
33. S. H. M. Nilar and S. Fraga, *J. Comp. Chem.*, **5**, 261 (1984).
34. S. Fraga, *J. Comp. Chem.*, **3**, 329 (1982).
35. E. Clementi and G. Corongiu, *Ions and Molecules in Solution*, N. Tanaka, H. Ohtaki, and R. Tamanushi, Eds., Studies in Physical and Theoretical Chemistry, Elsevier, Amsterdam, 1982, Vol. 27.
36. E. Clementi and G. Corongiu, *J. Biol. Phys.*, **11**, 33 (1983).
37. W. F. van Gunsteren, H. J. C. Berendsen, F. Colonna, D. Perahia, J. P. Hollenberg, and D. Lellouch, *J. Comp. Chem.*, **5**, 272 (1984).
38. J. M. Postma, H. J. C. Berendsen, and J. R. Haak, *Faraday Symp.*, **17**, 17/9 (1983).
39. Cray Research Inc., *Fortran (CFT) Reference Manual* (publication No. SR-0009), rev. 02, 1984, Part 3, pp. 2-1 to 2-16.
40. D. C. Rapaport and H. A. Scheraga, *Chem. Phys. Lett.*, **78**, 491 (1981).
41. W. F. van Gunsteren and M. Karplus, *Biochemistry*, **21**, 2259 (1982).
42. T. A. Andrea, W. C. Scope, and H. C. Andersen, *J. Chem. Phys.*, **79**, 4576 (1983).
43. M. Levitt and H. Meirovitch, *J. Mol. Biol.*, **168**, 595 (1983).
44. R. D. Swindoll and J. M. Haile, *J. Comp. Phys.*, **53**, 289 (1984).

W

A Molecular Dynamics Simulation of a Water Model with Intramolecular Degrees of Freedom.

Olle Teleman and Bo Jönsson
Physical Chemistry 2, Chemical Centre
POB 124 S-22100 Lund SWEDEN

Sven Engström
Physical Chemistry, University of Umeå
S-90187 Umeå Sweden

Abstract

Simulations of water models with and without intramolecular degrees of freedom have been performed within the classical approximation. Three H_2O simulations with interaction parameters taken from a simple point charge model due to Berendsen et al. are reported: i) with rigid water molecules, ii) with flexible water molecules and iii) with rigid water molecules and polarised bond lengths and bond angles obtained from the second simulation. The most important effect of the inclusion of intramolecular vibrations is seen on transport properties, which in general become a factor of 2-3 faster. The internal vibrations lead to a small but significant lowering of the potential energy and also reduce the structure, as seen from atom-atom distribution functions. The dipole moment fluctuations are affected by the nuclear polarisation caused by the intermolecular interactions. Two D_2O simulations have been performed in order to assess the effects of isotope substitution, one with rigid molecules and the other with flexible. Dynamics is somewhat slower for D_2O than for H_2O , whereas structural properties are unaffected. Changes are consistent with a kinematic coupling between intramolecular and intermolecular degrees of freedom.

Introduction

Most computer simulations of molecular liquids have been concerned with rigid molecules. This is particularly true for water, where the common models all treat the water molecule as a rigid body [1-5]. One exception is the central force model of Lemberg and Stillinger [6-9], which automatically includes the internal vibrations. This is achieved by treating oxygen and hydrogen atoms as separate particles but with an interatomic potential constructed so that H- and O-atoms form water molecules. Another approach has been made by Reimers and Watts [10-13], in which they combine a potential for the intermolecular interactions with a separate one handling intramolecular interactions. This technique but with different potential functions has also been applied by Bopp et al. [14].

The arguments for using a rigid water model seem to have both technical and physical origins. The technical argument against inclusion of internal vibrations is that a much smaller time step has to be used when integrating the equations of motion. This does not apply when using the multiple time step algorithm recently developed in our group [15]. Physically one can argue that internal vibrations are quantum mechanical in nature and can not strictly be incorporated into a classical Molecular Dynamics (MD) simulation. It remains to be shown, however, whether describing intramolecular vibrations classically, or treating them as rigid bonds is the best approximation to the true system.

Notwithstanding the above mentioned simulations of water models with internal degrees of freedom, no exhaustive comparison of dynamic and structural properties between flexible and rigid models exists to our knowledge. In this communication we report such a comparison of three MD simulations using the same site-site interaction model: i) with rigid bond lengths and bond angles, ii) with internal vibrations included and iii) as in the first case but with the average geometry derived from the second simulation. These simulations will be referred to as R1, F and R2, respectively. The difference between simulations R1 and R2 is that the water molecules will have a different geometry due to the nuclear polarisation obtained from the flexible simulation (F). In order to assess isotope effects two D_2O simulations were performed. The first will be referred to as the DR1 simulation, its parameters correspond to those of the R1 simulation (except for the deuterium mass). The second corresponds to the F simulation and will be designated by DF.

There are several advantages of using a flexible water model. One is that programming becomes simple and that it is easier to obtain a vectorizable code. It is also possible to calculate solvent induced changes in molecular properties and contributions to thermodynamic averages. Infrared shifts have been obtained in this way by Bopp et al. [14] and by Reimers and Watts [12,13]. The effect of nuclear polarisation on the dipole moment fluctuations may also be important for understanding the dielectric behaviour of aqueous solutions.

Anticipating some of the results, we finally advocate the use of dynamic properties when parametrising empirical or semi-empirical intermolecular potentials.

Simulation Technique

Trajectories were generated by numerical integration of Newton's equations of motion. The integration scheme was a predictor-corrector algorithm, where the predictor is a standard Taylor expansion and the corrector uses coefficients according to Gear [16].

In the rigid molecule simulations, a time step of 1.0 fs was used ($1 \text{ fs} = 10^{-15} \text{ s}$) and the angular motion was handled using quaternions [17]. The translational motion was integrated with a fourth order algorithm, while a third order algorithm was used for rotation. Doubling the time step leads to a rapid divergence within a small number of time steps.

In the flexible water molecule simulations (F and DF) all intramolecular degrees of freedom were treated explicitly. The oscillation period for the bond vibrational motion is of the order of 10 fs, and places an upper limit on the integration time step. To avoid any concomitant excessive demand on computer time, a multiple time step technique was used [15]. This amounts to a subdivision of forces into two components, one rapidly varying and one less so. The latter, which may then be evaluated less frequently, is also the more time-consuming and computing time reduces accordingly. In the present simulations, a short time step of 0.2 fs and a long time step of 1.2 fs were used. Increasing the long time step by a factor of two causes a very slow divergence, the symptom being an increase in intramolecular energy.

The system consisted of 216 water molecules enclosed in a cubic box with a side length of 18.6 Å. The temperature was calculated from the total kinetic energy. All simulations were performed under periodic boundary conditions and formally in the microcanonical ensemble. A molecular truncation of the interaction potential at 8.5 Å caused a numerical drift in the integration, amounting, in the F case, to 3.4% in total energy in 5000 time steps (1 ps). Temperature scaling was applied to remedy this. A neighbour list was used to keep track of interacting molecules, and it was updated every 4.8 fs.

The two MD programs have been developed in our laboratory. The program used for the simulation of flexible water molecules was run on a CRAY 1A and is designed to generate fast code on such a vector processor [15]. For the present system it takes 2.33 μs (total cpu time divided by number of interaction pairs) to compute the interaction between a pair of atoms and the whole trajectory of 57.6 ps took 2.58 CPU hours. The rigid water simulations were run on two different computers, an FPS-164 and on a Cyber 180/850. The latter two possess about the same computing capacity. On the Cyber 180/850 it took 33 μs to compute a pair interaction. Note that even so the program treating flexible molecules has to handle 50% more degrees of freedom. The large difference between the CRAY 1A and the Cyber 180/850 reflects the different general capacity of the two machines, but also that the flexible molecule MD program is more carefully adapted to take advantage of the Cray vector facilities. Both programs are completely general and handle all kinds of

mixtures. Technical details for the five simulations are summarized in Table I. Both programs are available on request.

Uncertainties given represent one standard deviation in the corresponding average. The uncertainties given for dynamical properties were obtained by dividing the sample into sub-systems, whereby averaging over these provided a basis for standard statistical analysis.

Molecular Potential

Of the many water models available in the literature we have chosen to use the Simple Point Charge (SPC) model of Berendsen et al. [3], the main reason being the simplicity of its functional form. The model contains point charges on hydrogen and oxygen, which interact according to classical electrostatics. The oxygen atoms also interact with a 6-12 Lennard-Jones term and the formal expression for the interaction between two sites i and j in two different molecules is

$$u_{ij}(r_{ij}) = q_i q_j / (4\pi\epsilon\epsilon_0 r) + 4\epsilon_{ij} [(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^6] \quad (1)$$

with ϵ set equal to 1. The covalent interactions within a water molecule are treated as harmonic oscillators [18],

$$u_{\text{intra}}(x) = (k_e/2) (x-x_e)^2 \quad (2)$$

where x is an instantaneous bond length or angle, x_e its corresponding equilibrium value and k_e the force constant. All potential parameters are given in Table II.

In the SPC model, the potential is assumed to be pairwise additive. Since it is parametrised from thermodynamic data, it is also an effective pair potential. The model thus includes, in an average way, any effects of intramolecular motion. In the present approach, molecular flexibility was added without adjustment of the non-covalent interaction parameters. Consequently, intramolecular flexibility is to some extent taken into consideration twice in the F and DF simulations. We will disregard this for present purposes, however, as our main purpose is not to compare with experimental data, but to investigate the consequences of introducing the internal vibrations explicitly.

Results and Discussion

The discussion of the present results for the rigid and flexible molecule approaches has to be qualitative in nature, since we do not know how important a quantum mechanical treatment of the internal vibrations is. This was also the reason for using a simple harmonic force field for internal degrees of freedom. More refined force fields including anharmonic corrections as well as coupling terms have been studied in several simulations [8,12-14]. While it has been shown [12,13] that inclusion of anharmonicity into the intramolecular potential is necessary in order to reproduce the water infrared spectrum satisfactorily, we only expect the anharmonicity to affect intermolecular couplings marginally. Unfortunately, there are no guidelines as to whether the classical approximation overestimates or underestimates the consequences of intramolecular flexibility. One may be inclined to regard the classical results as an underestimate thereof, since the quantum mechanical dispersion due to the zero-point energy, in both bond lengths and bond angle, is approximately twice as large as the classical [19]. Further effects exist, however, and the coupling between the different degrees of freedom will not be the same in a quantum mechanical model as in a classical. A full quantum mechanical treatment of all degrees of freedom in a simulation is still beyond reach, although recent approximate calculations have been published [19-23]. Another complication is the dependence of the dipole moment on molecular geometry. Here, it has been assumed to change according to geometry with fixed charges on the nuclei, which is a simplification [24].

Equilibrium Properties

The oxygen-oxygen, oxygen-hydrogen and hydrogen-hydrogen radial distribution functions are shown in Figure 1. For all three distributions, the flexible molecule (F) simulation displays significantly less structure than both the R1 and R2 simulations. In fact, there is little structure beyond the first peak in $g_{OO}(r)$. This contradicts a result of Reimers and Watts [13], who found an increased structure in all three radial distribution functions on introduction of internal vibrations. At present we do not see how to reconcile these observations. We note, however, that $g_{OO}(r)$ for the rigid model used in their work [13] lacks all secondary structure, whereas that of $g_{OO}(r)$ for rigid SPC water is comparable to that derived from experimental data.

A rather unexpected feature is seen when comparing $g_{OH}(r)$ and $g_{HH}(r)$ in Figures 1 and 2 for the R1 and R2 simulations. In the R2 simulation the height of the first peak and the depth of the first minimum are larger compared to those of the R1 simulation, but the remaining, long ranged, structure is about the same or even weakened. We interpret the first observation to be a consequence of the larger dipole moment in the R2 simulation. The higher multipole moments also vary with

molecular geometry. For instance, the $\mathbf{j}\mathbf{j}$ component of the quadrupole moment, where $\mathbf{j}$ is the hydrogen-hydrogen vector, decreases with decreasing bond angle. Such effects may account for the lessened secondary structure in the R2 simulation.

Kuharski and Rossky have made an entirely different attempt, treating intramolecular flexibility as a perturbation. They used classical normal mode distributions to stochastically modify the atomic positions in the trajectory from a Monte Carlo simulation of rigid water molecules. The perturbation to the individual molecules is uncorrelated to intermolecular position. Such a perturbation approach will reduce the structure in the radial distributions for purely mathematical reasons. The applied intramolecular perturbation can be averaged by means of the intermolecular orientation distributions to give the intermolecular perturbation. The latter will be a narrow function with a width of a few tenths of an Ångström. The final radial distribution $g(r)$ is now obtained as the convolution of the original $g(r)$ with the intermolecular perturbation. This perturbed $g(r)$ will thus possess less structure than the original one. Kuharski and Rossky found that $g_{\text{OH}}(r)$ and $g_{\text{HH}}(r)$ were modified by at most 3% and 5% respectively, in both cases losing structure. For completeness, we perturbed the R1 trajectory in two different ways and reanalysed it. Firstly, the bond length and bond angle distributions obtained from the F simulation were used for the perturbation. As these distributions are considerably narrower than corresponding distributions for the free molecule, the radial distributions of the thus perturbed R1 trajectory were altered by at most 2%. Secondly, we used the distributions for the free molecule for the individual bonds and angles. Then, $g_{\text{OH}}(r)$ and $g_{\text{HH}}(r)$ were modified by at most 3%. In other words, the perturbation approach does not reproduce the differences between the radial distributions of rigid water and those of flexible water, see Figs. 1 and 2. In view of these results it is obvious that correlation to intermolecular position is essential to the effects of intramolecular flexibility.

The average bond distance and bond angle found in the flexible molecule simulation change from their gas phase values. The changes are rather small, $+0.0163$ Å in bond length and -4.60° for the angle. This means that the average dipole moment in the F simulation is $8.14 \cdot 10^{-30}$ Cm (2.44 D) compared to $7.58 \cdot 10^{-30}$ Cm (2.27 D) for the original SPC water. Similar changes have been found by Reimers and Watts [13] from a Monte Carlo simulation using a more accurate intramolecular potential. The standard deviations of the bond length and bond angle distributions are given in Table I. The Boltzmann distributions obtained from the corresponding potential (cf. Eq. 2 and Table II) have standard deviations of 0.0328 Å and 6.56° for bonds and bond lengths respectively. The simulation distributions are thus narrower, especially for bond lengths.

The surprisingly large differences between the F and R2 simulations point to the difficulties met with when trying to treat intramolecular vibrations in an average way, as was attempted in the R2 simulation. These difficulties become even more apparent when considering the thermodynamic averages in Table III. For example, there is a large increase in pressure on inclusion of internal

vibrations. Other thermodynamic properties may be even more sensitive to the coupling between the intra- and intermolecular degrees of freedom.

In order to make a proper comparison of the total energy between simulations, an intramolecular energy of $3k_B T/2$ has been added to the values obtained from the R1 and R2 simulations. For flexible molecules $\langle U_{\text{intra}} \rangle$ has been defined as (cf. Eq. 2)

$$\langle U_{\text{intra}} \rangle = \langle k_e (x - x_e)^2 / 2 \rangle \quad (3).$$

With this definition, the intramolecular energy is found to increase by 1.4 kJ/mol from R1 to F. However, the total energy still decreases since the intermolecular energy decreases even more due to the nuclear polarisation. The net decrease in total potential energy is 2 kJ/mol, which is of the same order as the estimated non-additive contribution for the non-empirical MCY potential [2,25]. The intramolecular energy is unequally divided between the stretching and bending modes. Using Eq. 3, only $0.45 k_B T$ is found in each stretching mode while $1.1 k_B T$ is in the bending mode. These numbers may be sensitive to details in the intramolecular potential.

The dipole moment fluctuation density, calculated over a sphere, is shown in Fig. 3 as a function of the sphere radius. Unlike the properties discussed so far, the dipole moment fluctuation density is about the same in the F and R2 simulations. Without emphasizing the detailed shape of the curves, we see that the contribution from nuclear polarisation, i. e. the difference between the R1 and F curves, amounts to approximately 10%. This contribution may seem insignificant at first, but one should note that with the Clausius-Mosotti equation the dielectric permittivity ϵ is given by [26]

$$\langle m^2 \rangle \sim (\epsilon - 1) / (\epsilon + 2) \quad (4)$$

where m^2 is the dipole moment fluctuation density. If the dielectric permittivity is large, as it is for water, even a small change in the fluctuation density will lead to large changes in ϵ . A simple numerical calculation gives that a 10% reduction of $\langle m^2 \rangle$ corresponds to a decrease in ϵ from 80 to 21! In an empirical or semi-empirical potential this need not be a problem, but with non empirical potentials the use of a rigid model is probably as serious an approximation as is the assumption of pair-wise additivity. This is particularly so when studying long-range effects, which strongly depend on the dielectric screening.

Transport Properties

The diffusion coefficients and reorientational correlation times for the five simulations are summarized in Table IV. The general effect of the internal vibrations is to speed up the dynamics by a factor of two or three. Rotational motion seems to be slightly more affected than translational. The diffusion coefficient has been calculated from the mean square displacement as a function of time and the three curves are shown in Figure 4. The rotational motion has been studied by means of time correlation functions, $C(t)$, for three molecular vectors: One along the dipole moment, one in the plane of the molecule but perpendicular to the first and lastly a vector perpendicular to the other two. The first and second Legendre polynomials were evaluated:

$$C_1(t) = \langle P_1(t) \rangle = \langle n_i(t') \cdot n_i(t'+t) \rangle = \langle \cos \theta_i(t) \rangle \quad (5a)$$

$$C_2(t) = \langle P_2(t) \rangle = 0.5 \langle 3\cos^2\theta_i(t) - 1 \rangle \quad (5b)$$

where $\theta_i(t)$ is the angle between the vector $\underline{n}$ at time t' and $\underline{n}$ at time $t'+t$. The average runs over t' and all molecules present. The correlation times quoted in Table IV have been obtained from the long-time tails assuming exponential decay, cf. Figure 5. The ratio τ_1/τ_2 is significantly less than 3 for all vectors in all three models, which is less than one would expect from a diffusion model, where $\tau_1/\tau_2=3$, but in agreement with other water simulations [27, 28]. The internal vibrations largely affect the rotational motion around the different axes to the same extent, that is the relative ordering remains the same.

Isotope substitution

The transport properties of H_2O and D_2O are different. Experimentally, the translational diffusion is found to be $\approx 20\%$ slower in D_2O than in H_2O [29], and NMR measurements indicate that rotational correlation times are some 30% longer [30], both observations around room temperature.

From a simulation of rigid water molecules using an intermolecular potential due to Matsuoka et al. [2], the translational diffusion was found to be $\approx 30\%$ slower in D_2O than in H_2O [28]. This

was tentatively explained as due to a coupling between rotation and translation. However, in the same study Impey et al. [28] did not find any isotope effect on rotational correlation times.

The large changes seen on dynamic properties due to internal vibrations suggest that the experimentally found isotope shifts may well be explained by a coupling between intra- and intermolecular degrees of freedom. The results from a simulation of flexible D_2O molecules are given in Tables III and IV. The translational diffusion is unaffected by the isotope substitution, whereas the result for the rotational motion agrees with experiment. Thus, our results are completely contradictory to those of Impey et al. obtained for rigid molecules.

In a final attempt to understand these findings, we also performed a simulation of rigid D_2O molecules analogous to that of Impey et al.. Results are given in Tables III and IV. The isotope shift in the translational diffusion coefficient was approximately the same as that of Impey et al., whereas we also found an isotope shift in the rotational correlation times. We believe that this discrepancy can be explained by technical details, differing in their and our simulations.

We find that the experimental isotope shift is reproduced in the simulations of rigid molecules, both for translational and rotational motion. This is not so in simulations of flexible molecules. The issue seems complicated and further studies are needed before it will be fully understood.

Conclusions

Within a classical approximation we have found a coupling between inter- and intramolecular degrees of freedom. This coupling substantially affects most structural and thermodynamic properties. Proper treatment of internal vibrations seems essential in order to describe dielectric properties of polar liquids. The effect on transport properties appears to be even more important, increasing dynamics by a factor of two to three. The coupling to the intramolecular vibrations may also explain the isotope shifts found experimentally in translational and rotational diffusion coefficients.

References

- [1] Stillinger, F. H. and Rahman, A., 1974, *J. Chem. Phys.*, **60**, 1545.
- [2] Matsuoka, O., Clementi, E., and Yoshimine, M., 1976, *J. Chem. Phys.*, **64**, 2314.
- [3] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J., 1981, *Intermolecular Forces*, ed. B. Pullman, D. Reidel, Dordrecht.
- [4] Jorgensen, W. L., 1981, *J. Am. Chem. Soc.*, **103**, 335.
- [5] Jorgensen, W. L., 1982, *J. Chem. Phys.*, **77**, 4156. Jorgensen, W. L., 1983, *J. Chem. Phys.*, **79**, 926.
- [6] Stillinger, F. H., 1975, *Adv. Chem. Phys.*, **31**, 1.
- [7] Lemberg, H. L., and Stillinger, F. H., 1975, *J. Chem. Phys.*, **62**, 1677.
- [8] Rahman, A., Stillinger, F. H., and Lemberg, H. L., 1975, *J. Chem. Phys.*, **63**, 5223.
- [9] Stillinger, F. H., and Rahman, A., 1978, *J. Chem. Phys.*, **68**, 666.

- [10] Watts, R. O., 1977, *Chem. Phys.*, **26**, 367.
- [11] Reimers, J. R., and Watts, R. O., 1984, *Mol. Phys.*, **52**, 357.
- [12] Reimers, J. R., and Watts, R. O., 1984, *Chem. Phys.*, **85**, 83.
- [13] Reimers, J. R., and Watts, R. O., 1984, *Chem. Phys.*, **91**, 201.
- [14] Bopp, P., Jancso, G., and Heinzinger, K., 1983, *Chem. Phys. Lett.*, **98**, 129.
- [15] Teleman, O., and Jönsson, B., 1986, *J. Comp. Chem.*, in press.
- [16] Gear, C. W.: "Numerical Initial Value Problems in Ordinary Differential Equations"; Prentice Hall: Englewood Cliffs, N. J., 1971; pp. 148-154.
- [17] Evans, D. J. & Murad, S., 1977, *Mol. Phys.*, **34**, 327-331.
- [18] Lifson, S., 1982, *NATO Adv. Study Inst. Ser., Ser. A*, **45**, 359.
- [19] Wallqvist, A., and Berne, B. J., 1985, *Chem. Phys. Lett.*, **117**, 214.
- [20] Berens, P. H., Mackay, D. H. J., White, G. M., and Wilson, K. R., 1983, *J. Chem. Phys.*, **79**, 2375.
- [21] Kuharski, R. A., and Rossky, P. J., 1984, *Chem. Phys. Lett.*, **103**, 357.
- [22] Kuharski, R. A., and Rossky, P. J., 1985, *J. Chem. Phys.*, **82**, 5164.
- [23] Kuharski, R. A., and Rossky, P. J., 1985, *J. Chem. Phys.*, **82**, 5289.
- [24] John, I. G., Bacskay, G. B. & Hush, N. S., 1980, *Chem. Phys.*, **51**, 49-60.
- [25] Clementi, E., Kistenmacher, H., Kolos, W., and Romano, S., *Theor. Chim. Acta*, **55**, 257.
- [26] Neumann, M., 1983, *Mol. Phys.*, **50**, 841.
- [27] Rahman, A., and Stillinger, F. H., 1971, *J. Chem. Phys.*, **55**, 3336.
- [28] Impey, R. W., Madden, P. A., and McDonald, I. R., 1982, *Mol. Phys.*, **46**, 513.
- [29] Mills, R., *J. Phys. Chem.*, 1973, **77**, 685.
- [30] Lankhurst, D., Schrieffer, J. & Leyte, J. C., 1982, *Ber. Bunsenges. Phys. Chem.*, **86**, 215.

Tables

Table I. Technical details for the five simulations, two (R1 and R2) with rigid H₂O molecules, one with flexible (F), one with rigid D₂O molecules (DR1) and one with flexible (DF). The uncertainties given for bond lengths and bond angles are the standard deviation in the corresponding distribution.

Simulation	Bond length (Å)	Bond angle (°)	Temp (K)	Equilibration/ Production time
R1	1.0000	109.47	301±5	35.0 / 50.0 ps
F	1.0163±.0153	104.87±5.15	301±1	31.3 / 57.6 ps
R2	1.0138	104.62	301±5	35.0 / 50.0 ps
DR1	1.0000	109.47	301±6	35.0 / 50.0 ps
DF	1.0163±.0147	104.88±5.15	301±1	48.5 / 57.6 ps

Table II. Interaction parameters for the inter- and intramolecular potential functions.

q_O	=	-0.82 e	k_e^{OH}	=	4637 kJ·mol ⁻¹ ·Å ²
q_H	=	0.41 e	r_0^{OH}	=	1.0 Å
σ	=	0.3166	k_e^{HOH}	=	383 kJ·mol ⁻¹ ·rad ²
ϵ	=	0.65017 kJ·mol ⁻¹	α_e^{HOH}	=	109.47°

Table III. Thermodynamic averages from the simulations. Units are kbar and kJ/mol. The intramolecular energy is taken as $3k_B T/2$ for the rigid molecule simulations.

Simulation	< P >	< U _{inter} >	< U _{intra} >	< U _{tot} >
R1	0.7 ± 0.3	-41.8 ± 0.2	3.75	-38.0 ± 0.2
F	1.7 ± 0.2	-45.3 ± 0.3	5.14 ± 0.09	-40.1 ± 0.3
R2	0.0 ± 0.5	-47.6 ± 0.2	3.75	-43.8 ± 0.2
DR1	0.6 ± 0.6	-41.9 ± 0.2	3.75	-38.1 ± 0.2
DF	1.6 ± 0.2	-45.2 ± 0.3	5.02 ± 0.06	-40.2 ± 0.2

Table IV. Translational diffusion coefficients (D) and rotational correlation times (τ). Units are $10^{-9}\text{m}^2\text{s}^{-1}$ and 10^{-12}s . The three axes are a) the dipole vector, b) the H-H vector and c) a vector perpendicular to the molecular plane. The rotational correlation times given were derived from an exponential fit to the long time tail of the tcf:s, typically from 1-10 ps. The index on the correlation times refers to the order of the corresponding Legendre polynomial.

Simulation	D	$\tau_1(\text{a})$	$\tau_1(\text{b})$	$\tau_1(\text{c})$	$\tau_2(\text{a})$	$\tau_2(\text{b})$	$\tau_2(\text{c})$
R1	4.4 $\pm$ 0.1	3.13	2.69	1.99	1.33	1.66	0.97
F	6.1 $\pm$ 0.2	1.96	1.54	1.17	0.79	0.89	0.63
F2	2.6 $\pm$ 0.1	4.88 $\pm$ 0.11	4.44	3.23	2.17 $\pm$ 0.03	2.48	1.77
DR1	3.3	3.72	3.21	2.36	1.53	1.87	1.33
DF	5.9	2.40	1.69	1.48	1.00	1.10	0.85

Figure Captions

Figure 1. Atom-atom radial distribution functions. From top to bottom oxygen-oxygen, oxygen-hydrogen and hydrogen-hydrogen. Dashed lines represent the R1, dotted lines the R2 and solid lines the F simulation.

Figure 2. Difference plots for the atom-atom radial distribution functions of Figure 1. Solid lines represent $g^F(r) - g^{R1}(r)$ and dashed lines $g^{R2}(r) - g^{R1}(r)$.

Figure 3. Dipole moment fluctuation density as a function of the radius of the sphere over which the dipole moment was summed. Solid line refers to the F simulation, dashed line to the R1 simulation and dotted line to the R2 simulation.

Figure 4. Mean square displacement as a function of time for the oxygen atoms of the water molecules. Solid line refers to the F simulation, dashed line to the R1 simulation and dotted line to the R2 simulation.

Figure 5. Time correlation functions, cf. Eq. 5a, for the dipole vector of the water molecules (top) and their logarithms (bottom). Solid lines refer to the F simulation, dashed lines to the R1 simulation and dotted lines to the R2 simulation.

Fig. 1

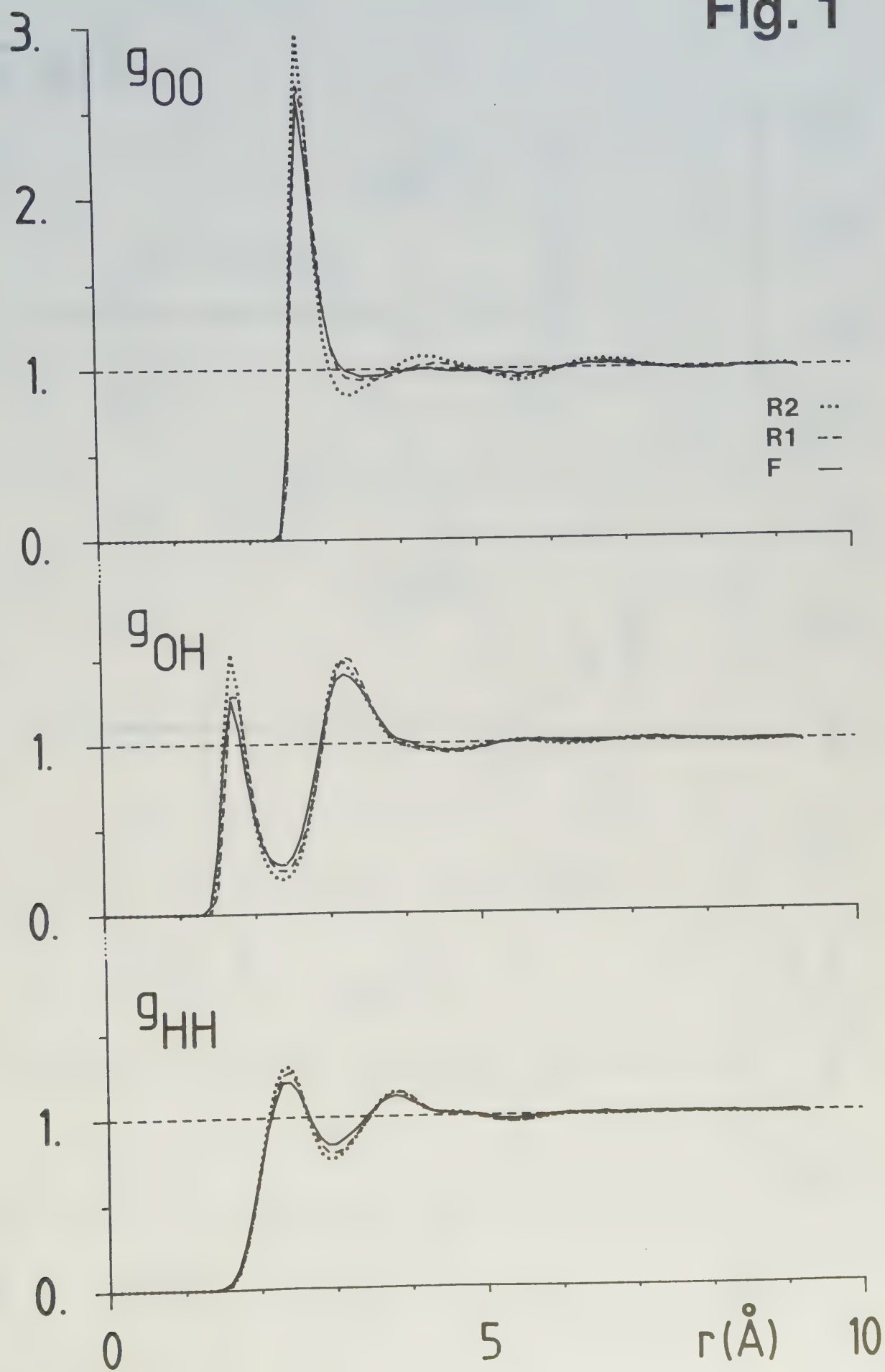
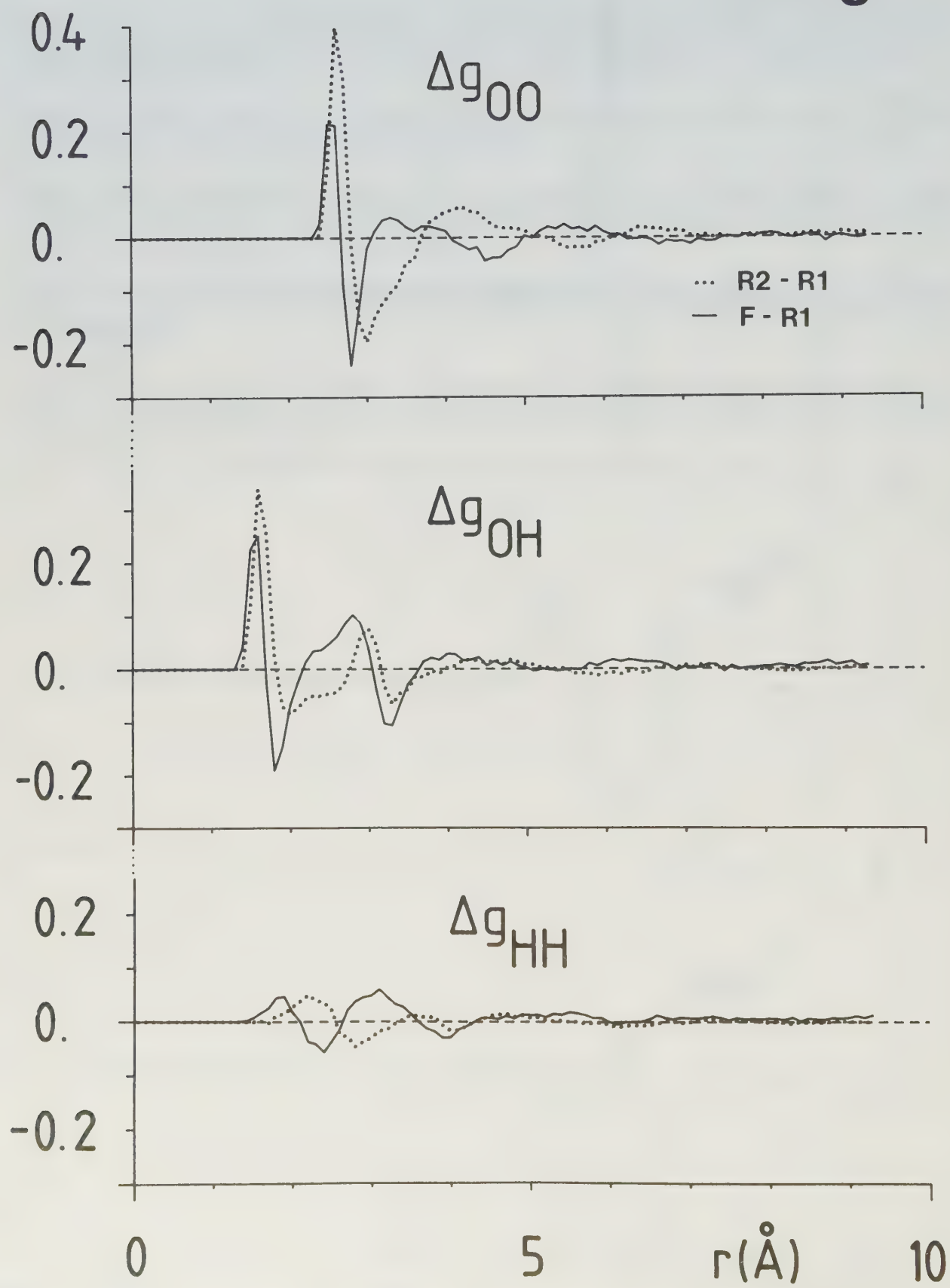


Fig. 2



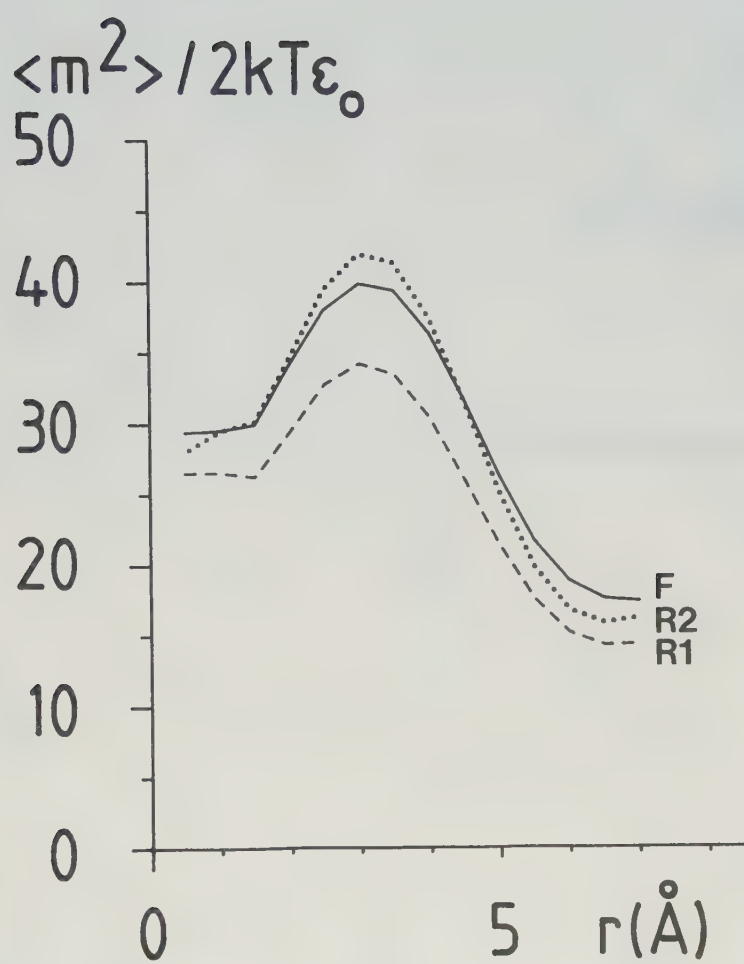


Fig. 3

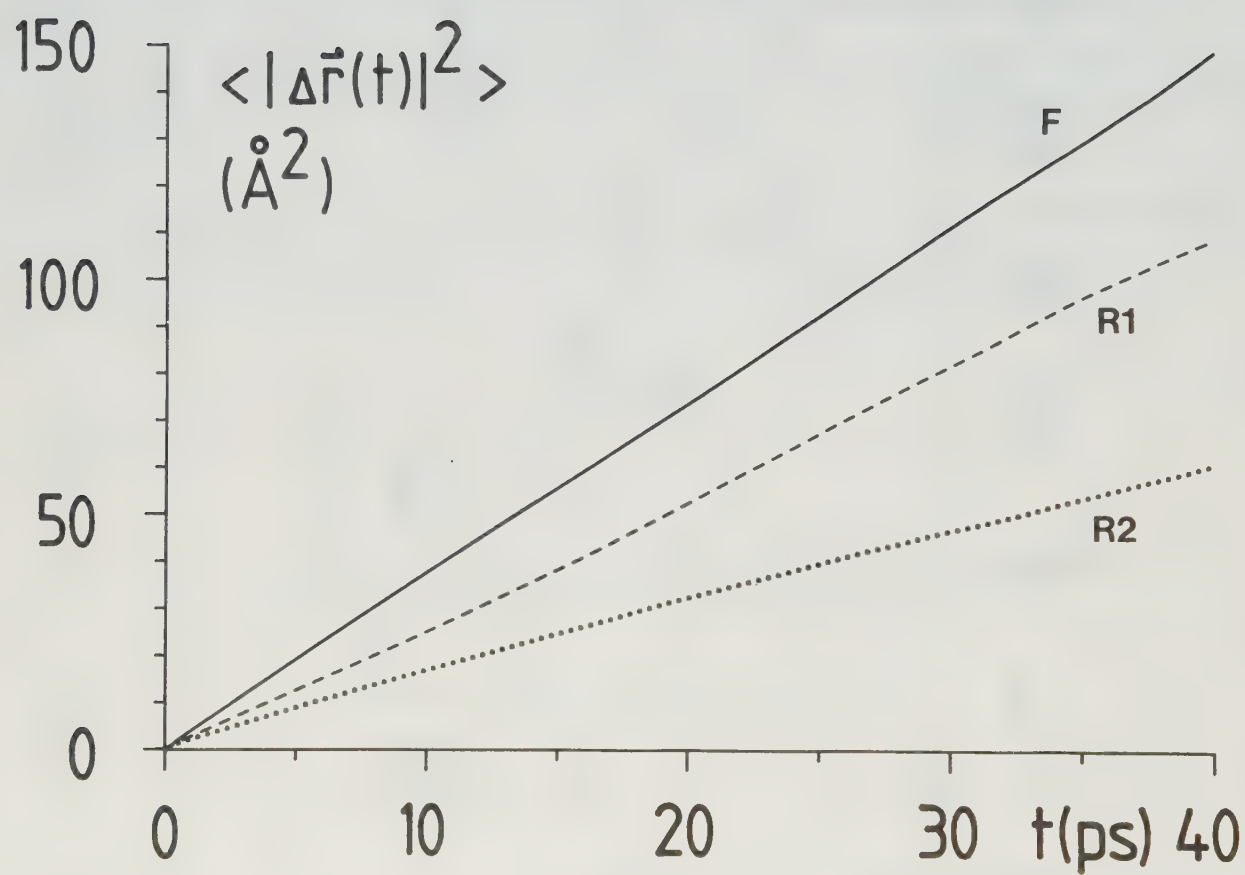
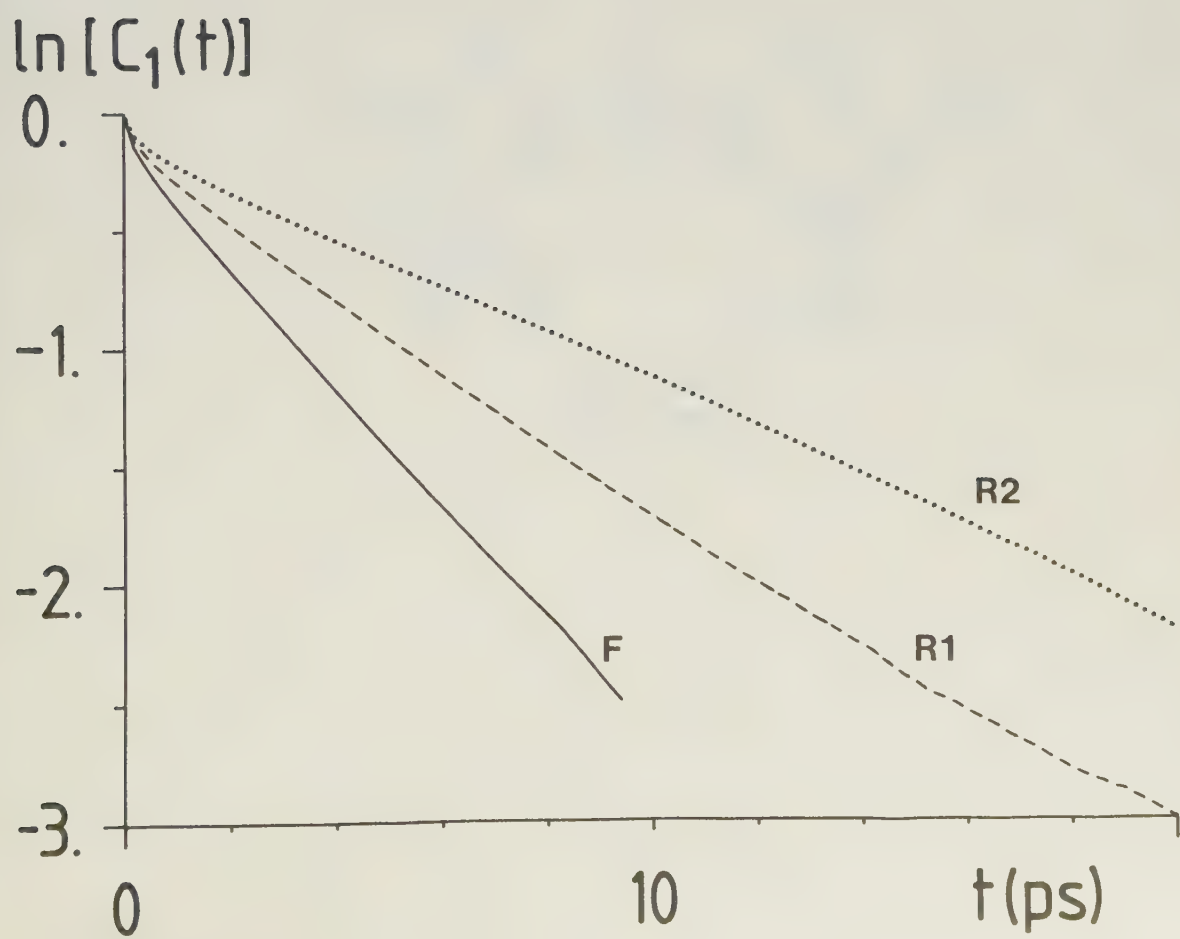
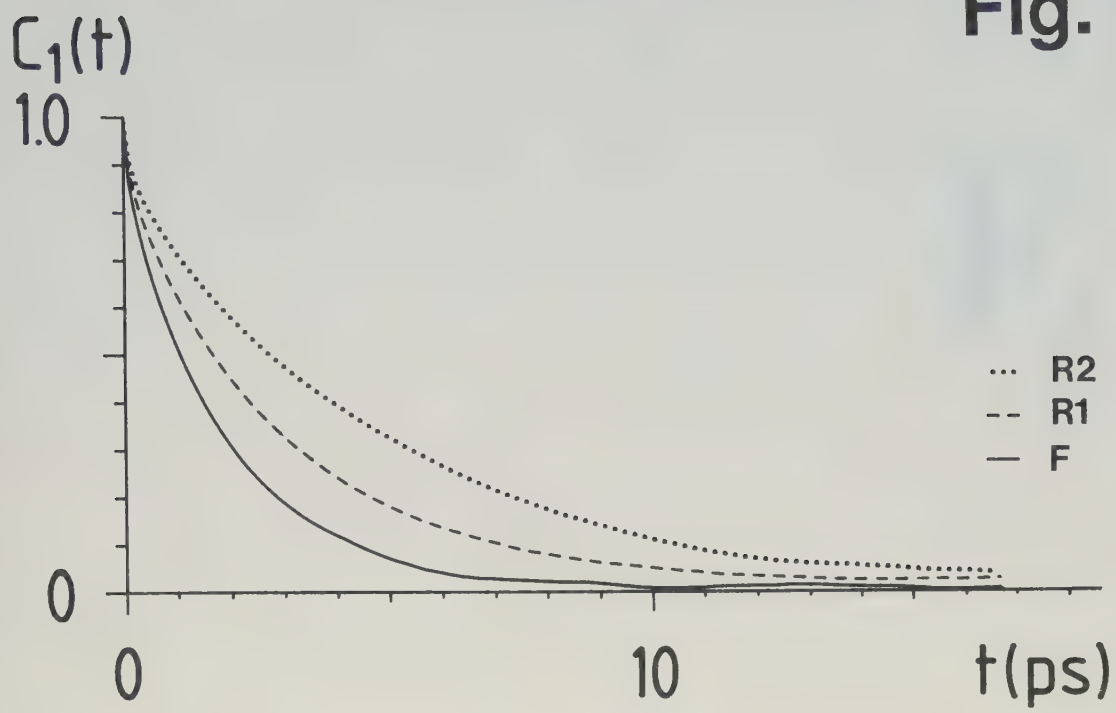
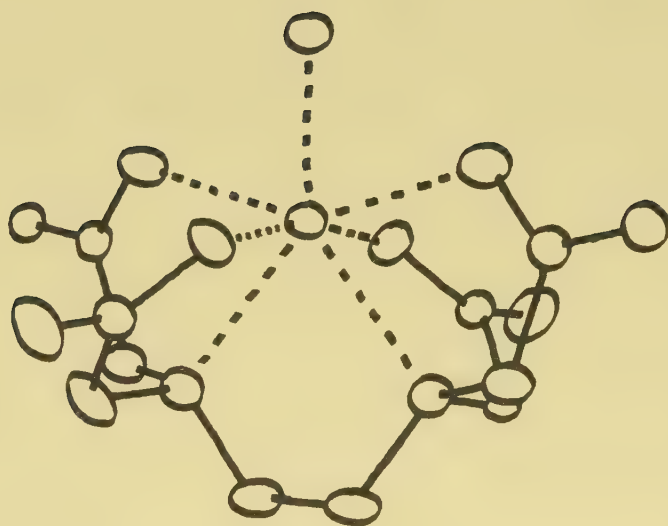


Fig. 4

Fig. 5



E



Molecular Dynamics Simulation of a Small Calcium Complex in Aqueous Solution

Olle Teleman & Peter Ahlström

**Dept. of Physical Chemistry 2, Chemical Centre,
University of Lund, POB 124, S-221 00 Lund, SWEDEN**

Abstract

A Molecular Dynamics technique is used to simulate CaEDTA^{2-} (two trajectories), H_4EDTA and a calcium ion in aqueous solutions, as well as pure water. The equation of motion is solved by means of a double time step algorithm. The potential function allows for full intramolecular flexibility. The structure of "flexible" water resembles that of corresponding rigid water models, while dynamics seem slightly faster. This also applies to all dynamics of molecules dissolved in flexible water. The simulated coordination number of Ca^{2+} in a 0.45 M solution is 7. Comparison of the two CaEDTA^{2-} trajectories indicates that quasi-ergodicity has been reached. In CaEDTA^{2-} the calcium coordination number is also ≈ 7 . Ligandation of calcium to the carboxyl groups is very strong, while the nitrogen-calcium relation is weaker. The structure of H_4EDTA is looser than that of CaEDTA^{2-} . Rotational correlation times for H_4EDTA are about half those of CaEDTA^{2-} , and the diffusion coefficient of H_4EDTA is about twice that of CaEDTA^{2-} . Both display some internal motion. Structural properties are plausibly reproduced, while dynamics seem fast. Results are compared to experimental data and discussed with respect to the properties of the potential.

Introduction

Binding of calcium ions to different ligands is an important process in nature, where calcium ions frequently act as intracellular transmitter substances¹. Many calcium binding proteins with varying functions and binding constants ($K=10^3$ - 10^8) are known¹. Some proteins, e. g. calmodulins² and troponins³, work as regulators and undergo considerable conformational change on calcium binding. They often display a strong cooperativity that increases their sensitivity to changes in the calcium concentration. In other proteins, e. g. α -lactalbumin, the calcium ion is virtually permanently bound and appears to be a strongly connective structural element. In all of these proteins the calcium ion binds preferably to oxygen atoms in different amino acid residues, above all aspartic and glutamic acid¹. The mechanisms of different processes involved in calcium binding is thus of great interest to molecular biology.

One way to obtain a deeper insight into the binding of calcium ions in proteins is to use computer simulation techniques. Recently we have embarked on such a study of the protein parvalbumin⁴. Such simulations, however, become rather complicated and very time-consuming. To provide a basis for their interpretation we have studied four smaller systems, namely pure water, one calcium ion in water, CaEDTA^{2-} in water and H_4EDTA in water. Pure water and calcium in water were mainly test systems for the model and the multiple time-step algorithm⁵. Comparison to experiment is possible since a wealth of X-ray and neutron diffraction data is available on the solvation of calcium ions, giving coordination numbers^{6,7} and mean residence times⁶ for the hydration shell of a calcium ion in water.

The CaEDTA^{2-} complex is suitable as a test system for calcium binding because experimental data on both structure and dynamics exists. The X-ray structure has been determined for both the $\text{Ca}(\text{CaEDTA})$ ⁸ and the $\text{Mg}(\text{MgEDTA})$ ⁹ complexes, thus providing two different starting configurations for a simulation. Diffusion coefficients¹⁰ as well as NMR relaxation data¹¹ for the CaEDTA complex are available and give a possibility to study the dynamic properties of the model chosen. The main chelating groups in EDTA are carboxylate groups similar to most calcium binding proteins. However, it is necessary to emphasize that the calcium binding of EDTA and of a protein differ in some respect, as in EDTA the interaction between the chelating groups and the calcium ion is the predominant energy contribution upon calcium binding, whereas in proteins the contribution to the potential energy from conformational changes may be equally important¹.

Simulation technique

Trajectories were generated by numerical integration of the Newtonian equation of motion, whereby all simulated systems were treated entirely as particle systems with only central forces. The integration was performed by a fourth order predictor-and-corrector algorithm, where the predictor is a standard Taylor expansion and the corrector uses coefficients according to Gear¹².

All intramolecular degrees of freedom were treated explicitly. The oscillation period of the faster vibrational modes amounts to $\approx 10^{-14}$ s and places a severe upper limit on the integration time step. To avoid any concomitant excessive demand on computer time a multiple time step technique was used⁵. This amounts to a subdivision of any force into two components, one rapidly varying and one less so. The latter, which may then be evaluated less frequently, is also the more time-consuming and computing time reduces accordingly. In all cases the small time step, ∂t , used was $0.2 \cdot 10^{-15}$ s. The large time step, ∂T , was $1.2 \cdot 10^{-15}$ s, except where otherwise stated.

The temperature was calculated from the total kinetic energy, which is a straightforward procedure, since no constraints limit the number of degrees of freedom. All simulations were performed under periodic boundary conditions and formally in the microcanonical ensemble. Truncation of the interaction potential caused a numerical drift in the integration, amounting to 3.4% in the total energy in 5000 time steps (1 ps) for pure water. To remedy this, temperature scaling was applied and the simulation ensemble is therefore in reality canonical. The pressure was calculated from the virial¹³:

$$p = \frac{NkT}{V} - \frac{1}{3V} \sum_{i=1}^N [\mathbf{r}_i \cdot \nabla_i U(\mathbf{r}_1, \dots, \mathbf{r}_N)] \quad (1)$$

where N is the number of atoms, V the volume, $U(\mathbf{r}_1, \dots, \mathbf{r}_N)$ the potential (see below) and ∇_i the gradient with respect to the position of atom i .

The MD program used has been developed in our laboratory and is designed to generate fast code on a Cray 1A vector processor⁵. Uncertainties given represent one standard deviation, except where otherwise specified.

Potential

The functional form of the potential is given by Eqs. 2 and 3. Eq. 2 contains the non-covalent contributions, which are assumed to consist of a 6-12 Lennard-Jones potential¹⁴ plus an electrostatic term;

$$U_{nc} = \begin{cases} \sum_{i < j} [4\epsilon_{ij} \{ \frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \} + \frac{q_i q_j}{4\pi\epsilon\epsilon_0 r_{ij}}] & , r_{ij} < r_{cut} \\ 0 & , r_{ij} > r_{cut} \end{cases} \quad (2)$$

where r_{ij} is the distance between two atoms i and j . Partial charges q_i and q_j were taken from the literature^{15,16}, and was -0.82, +0.41, +2.0, -0.635, +0.27, -0.06, +0.05 and -0.12 of an elementary charge for atom types 1-8. For numbering of atom types, see Fig. 1. For the Ca-EDTA complex the Lennard-Jones parameters ϵ_{ij} and σ_{ij} were calculated from the Kirkwood-Slater formula¹⁷ and from the assumption that the Lennard-Jones potential minimum corresponds to the sum of the van der Waals radii of the two atoms. Data were taken from Refs. 18-20. A large variety of water models is available in the literature, see ref. 20 and references therein. The simple point charge (SPC) model of Berendsen & al.²⁰ is parametrised according to Eq. 2 and was therefore used. The relative dielectric constant ϵ was assumed to be 1. Eq. 2 is taken to describe not only intermolecular interactions, but also those between atoms belonging to the same molecule, but separated by at least three covalent bonds. A large value for r_{cut} is necessary to account for the long-ranged electrostatic term of Eq. 2.

The covalent contributions to the potential (Eq. 3) comprise harmonic terms in all covalent bond lengths and bond angles, as well as periodic terms in all dihedral angles;

$$\begin{aligned}
 U_C = & \sum_{\text{bonds}} A_i (r_i - r_{i,e})^2 + \sum_{\text{angles}} B_j (\alpha_j - \alpha_{j,e})^2 + \\
 & + \sum_{\text{dihedrals}} [C_{k,0} + \sum_{l=1}^3 C_{k,l} \cos(l\beta_k)] \quad (3)
 \end{aligned}$$

where r_i is the actual length of the covalent bond i , $r_{i,e}$ the corresponding equilibrium length, α_j the actual value of the bond angle j , $\alpha_{j,e}$ the corresponding equilibrium value and β_k the actual value of the dihedral angle k . Parameters B_i , $r_{i,e}$, A_j , $\alpha_{j,e}$ and $C_{k,l}$ were taken from the literature¹⁹⁻²². While Reimers & Watts^{23,24} have shown that inclusion of anharmonicity into the intramolecular potential is necessary in order to reproduce the water infrared spectrum satisfactorily, we do not expect the anharmonicity terms to affect intermolecular couplings very dramatically. As this work is not aimed at a detailed investigation of high frequency vibrations, the potential parametrisation of Eq. 3 was assumed to be adequate. Also the hydrogen atoms within the EDTA molecule were treated explicitly. As no more than 12 nonpolar hydrogen atoms are present in any simulation, this

does not increase computing time significantly. All potential parameters used are given in Tables I-III.

The parameters of the SPC model describe effective interactions of rigid water molecules as obtained from a fit to thermodynamic data²⁰. The model thus includes, in an average way, any effects of intramolecular motion. In the present approach molecular flexibility was added without adjustment of the non covalent interaction parameters. In consequence they can not be relied on to represent an optimal fit to the above mentioned data. However, the inclusion of all intramolecular flexibility is computationally advantageous, since use of constraints is detrimental to the vectorisation properties of the program. It has not yet been clarified whether an exact quantum mechanical treatment is best approximated by a rigid, classical or by a flexible, classical model. This further motivates investigation of the latter approach.

Results and discussion

Pure water was simulated in the form of 216 water molecules enclosed in a cubic box with a side length of 18.6 Å. The cut-off distance for non covalent interactions (r_{cut}) was set to 8.5 Å. The initial system was a primitive cubic lattice with random molecular orientations. Simulation parameters are summarized in Table IV.

The temperature was set to 300 K. Temperature drift and scaling intervals were such that the average temperature increased to 301 ± 1 K. Thermodynamic simulation data are given in Table IV. The average total potential energy was -40.2 ± 0.2 kJ/mol, whereas for rigid SPC water the potential energy amounts to -42.2 kJ/mol²⁰. Our value includes, however, the potential energy of the three intramolecular degrees of freedom. Correction by $1.5kT$ ($=3.76$ kJ/mol) gives -43.9 kJ/mol. Reimers & Watts²³ give a value of -1.0 kJ/mole for the change in the water dimer potential minimum on inclusion of intramolecular flexibility. Many-particle effects and the inoptimality of potential parameters contribute to the remaining difference, and to the difference between our value and the experimental one (-41.64 kJ/mol, cf. Table III, p. 181 of Ref. 25). That the atomic charges in a molecule change on geometric rearrangement²⁶ is one example of neglected many-particle effects, others will be discussed below.

The bond length contribution to the potential energy, i. e. that arising from the first term of Eq. 3, was found to be $\approx 0.47kT$ per degree of freedom, very close to the equipartition value. The contribution to potential energy due to bond angles (second term of Eq. 3) amounts to $1.12kT$ per degree of freedom, which indicates that the angle bending is more strongly coupled to the intermolecular degrees of freedom.

The radial distribution function $g_{\text{OO}}(r)$ for the oxygen to oxygen distances is given in Fig. 2. It contains a nearest neighbour peak at 2.7 Å with a peak intensity of 2.6, as well as resolved second and third layers at 4.4 Å and ≈ 6.7 Å respectively, and thus largely resembles that of rigid

SPC water molecules²⁰. This applies also to the $g_{OH}(r)$ and $g_{HH}(r)$ radial distribution functions²⁵. All three $g(r)$ functions also resemble their counterparts obtained for the Matsuoka-Clementi-Yoshimine (MCY)²⁷ water model^{28,29}, whereas their secondary structure is somewhat less pronounced than for the two rigid water molecule models. Reimers & Watts²⁴ observed an increased structure in all three $g(r)$ functions on inclusion of intramolecular flexibility. We can not reconcile these data with ours as yet. However, it should be noted that $g_{OO}(r)$ for the rigid RWK2²⁴ model lacks all secondary structure, whereas that of $g_{OO}(r)$ for rigid SPC water is comparable to that derived from experimental data³⁰. This issue will be further addressed in a forthcoming article. Integration of the first peak in $g_{OO}(r)$ gives a number of nearest neighbours of about 4, integration of that of $g_{OH}(r)$ indicates that each water molecule engages in about four hydrogen bonds.

The diffusion coefficient, D , was calculated from Eq. 4,

$$D = \lim_{t \rightarrow \infty} \frac{1}{6t} \langle | \mathbf{r}_i(0) - \mathbf{r}_i(t) |^2 \rangle_i \quad (4)$$

where t is the time, $\mathbf{r}_i(t)$ the position of atom i and the average runs over all atoms. To obtain a measure of the uncertainty in D the trajectory was divided into three parts, each covering 19.2 ps, and a value of D calculated from each of them (D_1, D_2, D_3). The uncertainty, d , was then estimated by $|D_{\max} - D_{\min}|$, where $D_{\max}$ is the largest and $D_{\min}$ the smallest of the D_i 's ($i=1,2,3$). The diffusion coefficient was found to be $6.2 \pm 0.5 \cdot 10^{-9} \text{ m}^2/\text{s}$. While larger than the experimental value ($2.37 \cdot 10^{-9} \text{ m}^2/\text{s}$, see ref. 31), it is still inside the range of diffusion coefficients reported for rigid SPC water³². For pure water an uncertainty can also be obtained by dividing the system into subsystems and analysing these separately, thus providing a basis for standard statistical analysis. The uncertainty then amounts to $0.4 \cdot 10^{-9} \text{ m}^2/\text{s}$. This technique, however, is not applicable to systems that contain only one molecule of the investigated species, such as the EDTA systems, and we have chosen to use the former recipe throughout.

Reorientational autocorrelation functions for a vector $\vec{n}$ were calculated as:

$$C_j(\Delta) = \frac{1}{N} \sum_{i=1}^N \langle P_j(\cos\theta_i(t, \Delta)) \rangle_t \quad (5)$$

where $\theta_j(t, \Delta)$ is the angle between $\hat{n}(t+\Delta)$ and $\hat{n}(t)$, P_j is the Legendre polynomial of order j , the brackets indicate a time average and the summation runs over all water molecules. Below, the $C_j(\Delta)$ will be referred to as the j :th order time correlation function (tcf). Fig. 3 depicts the used vectors; one along the dipole moment of the water molecule, one in the plane of the molecule but perpendicular to the first and lastly one perpendicular to the other two. $C_2(\Delta)$ for these three vectors are shown in Fig. 4. These functions contain a rapid component, the decay of which is complete within ≈ 0.2 ps and which reduces the magnitude of $C_2(\Delta)$ by 30%-40%. This has been found in other simulations of liquid water and is usually interpreted as due to a hindered rotational motion of the water molecule²⁹. After 0.2 ps, $C_2(\Delta)$ can be very closely approximated with an exponential, which was obtained from a fit between 0.5 ps and 2.0 ps. The correlation time, τ_{2e} , corresponding to this exponential is given in Table V under the denomination "long times". The integral of the entire correlation functions defines the full correlation time, also given in Table V, and was calculated as (Δ in ps)

$$\tau_2 = \int_0^{0.5} C_2(\Delta) d\Delta + C_2(0.5) \cdot \tau_{2e} \quad (6)$$

where the integral from 0 ps to 0.5 ps was evaluated numerically. A comparison to MD simulation results on rigid SPC water²⁵ and MCY water²⁸ as well as with experimental data³³ is made in Table V. The reorientational correlation times of the flexible SPC water are smaller than those of rigid SPC water and rigid MCY water by approximately a factor of two. On the other hand, the differences between the vectors $\hat{n}_1$, $\hat{n}_2$ and $\hat{n}_3$ are qualitatively the same. Compared to experimental data³³, the correlation time, τ_2 , for $\hat{n}_2$ of the flexible SPC water is smaller by a factor of three.

The simulated diffusion coefficient and reorientation times are consistent, and describe a water molecule that is too mobile. It is clear that, while correctly representing the thermodynamic properties of water, it is less accurate in reproducing dynamic properties.

From experimental data on D ($2.37 \cdot 10^{-9} \text{ m}^2/\text{s}$, ref. 31) and τ_2 (2.06 ps, ref. 33) and the simulation results reported in this work, we note that these two properties scale in about the same way. It is thus tempting to see how well they can be accommodated within a hydrodynamic model. An effective radius may be calculated from the Stokes-Einstein formulae (Eqs. 7)³⁴:

$$D = \frac{kT}{6\pi\eta r_{e,D}}, \quad \tau_2 = \frac{8\pi\eta r_{e,\tau}^3}{6kT} \quad (7)$$

where k is the Boltzmann constant and η the viscosity. The effective radius, $r_{e,D}$ or $r_{e,\tau}$, is the radius of the hypothetical, spherical molecule that would have the same value of the dynamic entity (D or τ_2) as the actual one. With the experimental values of D , τ_2 and η ($0.865 \cdot 10^{-3} \text{ kg/ms}$) one finds $r_{e,D} = 1.08 \text{ \AA}$ and $r_{e,\tau} = 1.33 \text{ \AA}$. The latter is larger by $\approx 24\%$, giving a measure of the accuracy with which the hydrodynamic model can be considered to describe real water. If we assume that $r_{e,D} = r_{e,\tau}$ we can insert the simulated values of D and τ_2 into Eqs. (7) in order to obtain an effective radius:

$$r_{\text{eff}} = 3 [0.5\tau_2 D]^{0.5} \quad (8)$$

and a viscosity:

$$\eta = \frac{kT}{18\pi D} [0.5\tau_2 D]^{-0.5} \quad (9)$$

Equations (8) and (9) now yield $r_{\text{eff}} = 1.3 \pm 0.4 \text{ \AA}$ and $\eta = 0.27 \pm 0.07 \cdot 10^{-3} \text{ kg/ms}$ respectively.

One calcium ion in water. In order to obtain a picture of the hydration properties of a calcium ion as described by our potential, one Ca^{2+} ion and 124 water molecules were simulated. The system was enclosed in a cubic box with a side length of $15.5 \text{ \AA}$. Initially, the calcium was placed at the centre of the box and the water molecules were centered on the bases of a primitive cubic lattice, with their orientations randomized. Simulation parameters are given in Table IV.

The large time step was $2 \cdot 10^{-15} \text{ s}$. The temperature was set to 300 K, and the average temperature was 311.3 K. The short cut-off radius and the ionic nature of the system cause this large temperature drift. Thermodynamic averages are given in Table IV.

The radial distribution function $g_{\text{Ca-O}}(r)$ for the distances between the calcium ion and all water oxygens is shown in Fig. 5 and contains two well resolved neighbour layers. That of nearest neighbours is centered at $2.30 \pm 0.14 \text{ \AA}$ and integration yields a coordination number of 7.0 ± 0.2 .

Probst & al.⁷ have obtained a value of ≈ 9 for the calcium ion coordination number from MD simulations of CaCl_2 in water using a three-body, central force potential^{35,36}. From X-ray diffraction data, however, they find a value of ≈ 7 , in agreement with our results. The next layer contains about 20 water molecules at an average distance of 4.8 Å. The nearest neighbour peak in $g_{\text{Ca-H}}(r)$ contains 14.3 ± 0.4 hydrogens and is situated at 3.0 ± 0.2 Å. In order to characterize the structure of the first hydration shell, we have employed the angular probability density $Q(\theta, \phi)$, where θ and ϕ are the polar angles of the water dipole vector with respect to an axis defined by the vector from the calcium ion to the oxygen atom of the water molecule. Due to symmetry we can integrate over the azimuth, ϕ , to obtain $Q(\theta)$. The distribution of θ angles, $P(\theta)$ is now obtained as $Q(\theta)$ multiplied by the volume element, i. e. proportional to $Q(\theta) \cdot \sin\theta$. One finds an average of $\langle\theta\rangle = 20^\circ \pm 11^\circ$, whereas $Q(\theta)$ has a narrow maximum at 0° . The orientation distribution of liganded waters is thus, naturally enough, strongly polarised by the calcium ion. From neutron diffraction measurements, Hewish et. al.³⁷ obtain a mean angle of $38^\circ \pm 9^\circ$ for a 1 molal CaCl_2 solution. The Ca^{2+} coordination number is given as 10.0 ± 0.6 , which deviates by 3 from our findings. They also find the coordination number to be strongly increasing with decreasing concentration. As the simulated solution had a molality of 0.45, this increases the difference in coordination number, and also makes a direct comparison to our results on the tilt angle difficult. Data are available also on NiCl_2 solutions³⁸, and a mean tilt angle of $17^\circ \pm 10^\circ$ is given for a 0.46 molal solution, where also the cation coordination number, reported to be 6.8 ± 0.8 , is very close to our simulation value.

The diffusion coefficients of the calcium ion was obtained from the slope of the mean square displacement, and was found to be $2.7 \pm 0.4 \cdot 10^{-9} \text{ m}^2/\text{s}$. Mateo & al.¹⁰ give an experimental value of $0.91 \cdot 10^{-9} \text{ m}^2/\text{s}$ at 308 K, which corresponds to $0.97 \cdot 10^{-9} \text{ m}^2/\text{s}$ at 311 K. Our simulation value is larger by a factor of ≈ 2.8 , which is very close to the ratio between simulation and experimental values of the water self-diffusion coefficient. The correlation function $C_2(\Delta)$, cf. Eq. 5, was calculated for a vector from the calcium ion to the oxygen of its water ligands. The corresponding correlation time, $\tau_{2,\text{lig}}$, was 1.9 ps and is probably mainly determined by diffusion of liganded water molecules within the hydration shell. The preferred direction of the dipole vectors of liganded water molecules is manifested also in intramolecular correlation times. The τ_2 of the dipole vectors is 1.7 ps ($\hat{n}_1$, see Fig. 3) and thus nearly equals $\tau_{2,\text{lig}}$. On the other hand, τ_2 for $\hat{n}_2$ and $\hat{n}_3$ amount to 1.0 ps and 0.6 ps respectively, and are thus comparable to the values for bulk water molecules, cf. Table V.

During the trajectory five ligand exchanges occur, which corresponds to a mean ligand residence time, τ_M , of ≈ 50 -60 ps, where τ_M is defined as the mean time a particular water molecule lives in the complex before changing place with a water molecule outside it. This number is very uncertain for statistical reasons and is larger than those obtained for some monovalent ions from MD simulations on electrolytes, using the MCY water potential³⁹. One author⁶ gives an upper limit of 10 ps for the residence time at normal pressure (≈ 1 bar). For the ligand exchanges in our trajectory, the departing ligand leaves before the arriving one is inserted, in four cases out of five. Thus, a large pressure would seemingly induce an increase in residence time, which may be the cause of the simulation value of τ_M , as the pressure was ≈ 2000 bar.

EDTA simulations performed. Two simulations of CaEDTA in water were performed. One used crystal coordinates of $\text{Ca}(\text{CaEDTA}) \cdot 7\text{H}_2\text{O}$ as starting coordinates⁸. One of the two calcium ions occupies the EDTA binding site and has only two water ligands. The coordinates of these four molecules were used, and the remaining calcium ion and five water molecules were discarded. This complex was immersed in a further 312 water molecules in a cubic box with a side length of 21.7 Å. The water molecules, the orientations of which were randomized, were placed on the bases of a primitive cubic lattice, except for the volume occupied by EDTA. Below, this simulation will be referred to as the "C" simulation. Following on this, the Ca^{2+} ion was removed and the four carboxyl groups were covalently protonated, which corresponds to a low pH value. The system was then left to reequilibrate, whereafter a further trajectory was recorded. This will be referred to as the "H" simulation in the following.

The second simulation was performed to obtain a check on equilibration times and on the sensitivity to starting configuration. As initial configuration, the coordinates of MgEDTA^{2-} and those of the single magnesium liganded water molecule from the crystal coordinates⁹ of $\text{Mg}(\text{MgEDTA}) \cdot 9\text{H}_2\text{O}$ were used. The Mg^{2+} was replaced by a Ca^{2+} and 316 water molecules were added. The simulation will be referred to as the "M" simulation. Parameters for the three trajectories are given in Table IV.

Thermodynamic averages for these simulations are given in Table IV. The difference in potential energy between the "H" simulation on one hand and the "C" and "M" simulations on the other is of course due to the absence of the strong calcium-ligand electrostatic interactions in the "H" case. Using the EDTA crystal coordinates one obtains that the electrostatic energy of CaEDTA^{2-} is lower than that of H_4EDTA by ≈ 2500 kJ/mol. This figure is comparable to the simulation results.

Structural properties of EDTA. In no case ("C" or "M") is the calcium ion released from the complex. However, we do observe changes in ligandation. In the "C" trajectory one carboxyl ligand is released and in two of the other carboxyl groups, ligandation is transferred from

one oxygen atom to the other. One water ligand is added. In the "M" trajectory, one carboxyl group turns to make both oxygen atoms interact strongly with the calcium. One water ligand is added and the original water ligand is exchanged.

Radial distribution functions for the carboxyl oxygen atoms, the nitrogen atoms and the water oxygen atoms, all with respect to the calcium ion, are plotted in Fig. 6 for the "M" simulation. The corresponding functions for the "C" case are very similar, as indicated by Fig. 7 (see below), except for the occurrence of distances larger than 4 Å in the "Ca²⁺ to EDTA oxygens" function. This is a consequence of the above mentioned carboxyl ligand release. In both the "C" and the "M" trajectories the two nitrogen atoms slightly recede from the calcium ion, to a distance of 3.2 ± 0.3 Å, and in the "C" case, where one carboxyl ligand was released, even farther away. The carboxyl ligands remain much closer to the calcium ion, i. e. at an average distance of 2.3 ± 0.1 Å. The integral of the first peak amounts to ≈ 4.6 oxygen atoms and the remaining ≈ 3.4 are not liganded to the calcium ion (average distance 4.2 Å). The interaction between carboxyl groups and their spatial neighbours reduces the average carboxyl O-C-O angle from its vacuum equilibrium value of 129° to $\approx 121^\circ \pm 5^\circ$, whereas the average C-O bond length is not significantly altered. The first moment of the carboxyl group charge distribution thereby increases by $\approx 14\%$. For the relatively few cases where both oxygens ligand the calcium ion, the angle is reduced to $116^\circ \pm 5^\circ$, further increasing the polarity of the group. Three peaks are well resolved in the water oxygen radial distribution functions. The nearest neighbour peak, which contains the liganded water molecules, is situated at 2.3 ± 0.1 Å, and amounts to ≈ 2.0 water molecules (this number increases to about 3 when, in the "C" trajectory, one carboxyl ligand has been released). The next peak is centered at 4.3 ± 0.5 Å and contains about 13-16 water molecules. This shell consists of the water molecules that are hydrogen bonded to the actual calcium ligands, but is also partly due to the general shape of the EDTA complex. There is also a third, weakly pronounced water shell at ≈ 6.6 Å, whereas outside ≈ 7.5 Å the constant bulk water density prevails.

The sum of these three radial distributions is plotted in Fig. 7 for the "C" and "M" simulations. The similarity of the two functions is obvious. All candidates for calcium ligandation are comprised in this "total" rdf, thus the number of ligands can be obtained from its integral. Conclusion of the integration at 2.7 Å gives a coordination number of ≈ 6.7 , whereas extending the integration to 3.5 Å will include the two nitrogen atoms to make a total of ≈ 8.4 ligands. Note that in the Mg(MgEDTA)·9H₂O crystal complex the central Mg²⁺ ion (replaced by a Ca²⁺ in simulation) is heptacoordinated, whereas in the crystal Ca(CaEDTA)·7H₂O the central ion possesses eight ligands. In comparison, a luminescence study⁴⁰ of EuEDTA⁻ and TbEDTA⁻ in aqueous solution showed that Eu³⁺ and Tb³⁺ in these complexes have about three water ligands. As europium and terbium normally have slightly larger coordination numbers than calcium, this finding is consistent with our simulation results on CaEDTA²⁻.

Comparison of the general structure of the CaEDTA^{2-} complex in different environments was achieved by means of R factors, as applied in the refinement of X-ray crystallography data⁴¹. The R factor for a pair of configurations a and b, denoted by $R(a,b)$, was taken to be the minimum of the function R of Eq. 10:

$$R = \left[\sum_{i=1}^N m_i |\mathbf{r}_{i,a} - \mathbf{r}_{i,b}|^2 / \sum_{i=1}^N m_i \right]^{0.5} \quad (10)$$

where m_i is the mass of atom i and $\mathbf{r}_{i,a}$ and $\mathbf{r}_{i,b}$ its position in the two configurations. The summation covers Ca^{2+} and the atoms of EDTA^{4-} , in all 33 atoms, except for where one of the configurations is taken from the "H" simulation, when the R factor was calculated on the basis of the 32 EDTA^{4-} atoms. Minimization is performed with respect to interconfigurational position and orientation. The R factor provides a comprehensive, quantitative measure of the similarity of configurations. Further, the minimization of R provides the translation and rotation for optimal superposition, and thus the basis of any comparison of specific parts of the structure. To obtain an indication of the upper limit of R factors, the EDTA configuration at the starting point of the "M" simulation was taken and an entire half of the molecule was rotated 180° around the C-C bond of the central ethylene, cf. Fig. 1. The R factor between this configuration and the one before rotation was the evaluated and found to be 2.32 Å. This value is thus to be considered very large for EDTA.

Some R factors are shown in Table VI. We will use the following notation to facilitate the discussion. Let XM denote the MgEDTA crystal coordinate set, XC the CaEDTA crystal coordinate set, C_{bef} a configuration from the "C" trajectory taken before the release of the carboxyl ligand, C_{aft} one taken after that release, and M and H configurations from the "M" and "H" trajectories respectively. We note that $R(C_{\text{bef}}, XC) \approx R(M, XC)$ and that $R(C_{\text{bef}}, XM) \approx R(M, XM)$. Further, both trajectories display CaEDTA^{2-} configurations that on average are somewhat closer to the crystal MgEDTA^{2-} structure. This is most likely due to the fact that the second metal ion is in part liganded directly to the EDTA molecule in the $\text{Ca}(\text{CaEDTA}) \cdot 7\text{H}_2\text{O}$ crystal, and thus distorts it, which is not the case for the $\text{Mg}(\text{MgEDTA}) \cdot 9\text{H}_2\text{O}$ crystal. This is also reflected by $R(XC, XM)$ being as much as 1.36 Å. $R(C_{\text{bef}}, M)$ approximately equals $R(C_{\text{bef}}, XM)$ and $R(M, XM)$, while $R(C_{\text{aft}}, M)$ is close to $R(C_{\text{aft}}, XM)$. However, $R(C_{\text{bef}}, M)$ features some very low minima, the lowest of which is 0.33 Å, indicating very similar conformations. Thus there is at least one area of configurational space, that both trajectories have traversed, albeit originating from very different starting points ($R(XC, XM) = 1.36$ Å). For the "H" trajectory, we note that $R(H, d)$ is not larger than $R(C_{\text{aft}}, d)$, where d belongs to the set $D = \{XC, XM, M\}$, i. e. that the "H" configurations do not deviate more

from the configurations of set D than do the configurations where one EDTA calcium ligand has been replaced by water. We may reword that as follows. If we take XM or M as a reference set, we find that the configurations where a water molecule has supplanted a carboxyl ligand, forcing the latter outwards, are deformed to about the same extent as those where the four carboxyl groups may move freely. Note also that the carboxyl groups each have a hydrogen in the "M" case and thus do not repel each other electrostatically. On the other hand, $R(H,b)$ is significantly larger than $R(a,b)$, where a and b belong to the set $A=\{XC, XM, C_{bef}, M\}$, thus "H" configurations deviate more from all configurations of set A than any from that set deviates from any other. In addition, no configuration pairs H, C_{bef} nor H,M could be found possessing an R factor below 0.83 Å, which is very much larger than the minima for C_{bef} -M. In other words, there is no H configuration that resembles any C_{bef} configuration, and none that resembles any M configuration. We conclude that the simulations very clearly distinguish between the EDTA where the calcium ion is present, and that where it is not.

The similarities between the "M" and "C" trajectories discussed here and above indicate that the structural properties of the system are determined by the potential and the macroscopic parameters, i. e. that the system is quasi-ergodic, and also that an equilibration time of ≈ 17 ps is adequate for this kind of system.

Dynamic properties of EDTA. First and second order reorientational correlation functions (cf. Eq. 5) were calculated for a set of vectors of the $CaEDTA^{2-}$ complex. Two of these are shown in Fig. 8. First estimates of the corresponding correlation times were obtained by fitting an exponential between 0.5 ps and 6.0 ps to the tcf:s. For such an estimate to be accurate, it is necessary that the length of the simulation be very much larger than the estimate itself, which is not the case here. Moreover, the tcf:s themselves possess but little resemblance to singly exponential functions, and in consequence the general definition of the correlation time as the integral of the correlation function (normalized so as to make $C(0)=1$) is to be preferred. The use of this definition is also prerequisite of any comparison to NMR data. Straightforward evaluation of that integral is not possible, however, as the tcf:s are available only up to some time Δ_L and their tail for times larger than that normally contributes considerably to the integral. In principle, Δ_L equals the length of the trajectory, T_L , but the statistic uncertainty is acceptably small only at times $\Delta \ll T_L$, and thus $\Delta_L \ll T_L$. To circumvent this difficulty one may approximate the tail of the correlation function by $C(\Delta_L) \cdot \exp[-(\Delta - \Delta_L)/\tau]$. Then

$$\begin{aligned}
 \tau &= \int_0^{\infty} C(\Delta) d\Delta = \int_0^{\Delta_L} C(\Delta) d\Delta + \int_{\Delta_L}^{\infty} C(\Delta_L) \exp[-(\Delta-\Delta_L)/\tau] d\Delta = I + \tau C(\Delta_L) \\
 &\Leftrightarrow \\
 \tau &= I / [1 - C(\Delta_L)]
 \end{aligned} \tag{11}$$

where I is the (numerically evaluated) integral of $C(\Delta)$ from 0 to Δ_L . If $C(\Delta)$ contains a well discernible and rapidly decaying component ($\tau_f \ll \Delta_L$), this may be taken into account. Take

$C(\Delta) = AC_f(\Delta) + BC_s(\Delta)$ with $A+B=1$, and A and B known. Then $\tau = A\tau_f + B\tau_s$. Now, if $A\tau_f \ll \tau$ one obtains

$$\begin{aligned}
 \tau &= \int_0^{\Delta_L} C(\Delta) d\Delta + \int_{\Delta_L}^{\infty} C(\Delta_L) \exp[-(\Delta-\Delta_L)/\tau_s] d\Delta \approx I + \tau C(\Delta_L)/B \\
 &\Leftrightarrow \\
 \tau &= BI / B - C(\Delta_L)
 \end{aligned} \tag{12}$$

Correlation times were obtained from Eq. 12 and are given in Table VII. For the C-H vectors the average tcf for all eight vectors was analysed. The uncertainties in the correlation times are the standard deviation in averaging the different correlation times obtained as a result of using different values of Δ_L , evenly distributed from 3 ps to 24 ps. These uncertainties are very much less than the range of correlation times obtained for individual methylene groups, also given in Table VII. For reasons of symmetry, all four acetic acid methylenes should have the same correlation times, but only in the limit $T_L \rightarrow \infty$, a condition which is obviously not fulfilled in these simulations. Given in Table VII is also an experimental value of the C-H second order correlation time as obtained by ^{13}C NMR¹¹. The ratio to the simulation value is about 7.

The mean square displacement of the EDTA complex center of mass was calculated as a function of time, from which the diffusion coefficient was obtained, cf. Table VII. From experimental data, Mateo & al.¹⁰ give a value of $\approx 0.59 \cdot 10^{-9} \text{ m}^2/\text{s}$ at 303 K, which is about half the simulation value.

Discussion of EDTA properties. There are several difficulties in simulations of large molecular systems, particularly as regards comparison to experiment. One such is the intermolecular potential, which here, as in most other simulations reported, has been assumed to be pairwise additive. This is obviously an approximation, especially for charged systems, which

partly can be remedied by using an effective pair potential, which includes the nonadditive terms in an average way. It is, however, very difficult to assess the accuracy of this approximation. Thus we can not *a priori* expect MD simulations to achieve full reproduction of physical properties. The strong interaction of the calcium ion and the nitrogen atoms of the EDTA molecule in the crystal structures is mostly due to the nitrogen lone pair of electrons, which is strongly polarised by the electric field of the calcium ion. We interpret the weak ligandation of the nitrogen atoms in the simulations to be caused by the absence of this term in the potential. Concomitant to this partial neglect of the inductive term in the calcium-nitrogen relation is an increase in internal mobility.

Also the carboxyl groups will be polarised by the calcium ion, and the effective charge transfer across the carboxyl group will be significant. A charge transfer of 0.1 e amounts to a decrease in potential energy of ≈ 60 kJ/mol per carboxyl group. The polarisation will strengthen the ligandation of the carboxyl group as a whole. As the polarisability of the carboxyl group is larger in the direction from one oxygen atom to the other, the polarisation will also increase the energy barrier for rotation of that group. From the energies involved, this increase in barrier height will decrease the probability of carboxyl flips, and even more so the probability of a carboxyl ligand release, by several orders of magnitude, and they might be entirely inhibited on the time scales accessible to MD simulation. This line of reasoning is consistent with the very low rate constant found for the dissociation of Ca^{2+} from EDTA⁴². The observed carboxyl flips (2 in the "C" trajectory, 1 in the "M" trajectory), and the carboxyl group release ("C" trajectory), certainly entail increased motion in adjacent parts of the complex. This internal motion is substantiated by the correlation times being shorter for the acetic acid C-H vectors than for the N-N and Ca-N vectors, where the tcf:s of the latter may be assumed to represent reorientation of the entire complex.

There is some evidence in the literature that EDTA chelates calcium in two forms, either hexadentately or pentadentately⁴³. In the latter case one carboxyl oxygen ligand has been replaced by a water molecule, as occurred in the "C" trajectory. The experiments were, however, interpreted on the assumption that Ca^{2+} has a coordination number of 6, and it is not clear to us whether a larger coordination number would affect this interpretation or not.

The presence of internal motion leads to a partial breakdown of the hydrodynamic model. While the diffusion coefficient more or less monitors the bodily size of the CaEDTA^{2-} complex, the reduction in correlation times caused by intramolecular flexibility precludes consistency within the model as formulated by Eqs. 7. From the simulation value of the water viscosity, the EDTA diffusion coefficient and second order reorientational correlation time, Eqs. 7 yield $r_{e,D}=7.4$ Å and $r_{e,\tau}=3.8$ Å, values that differ by nearly a factor of 2. In comparison, the volumes of the crystal Ca^{2+} and Mg^{2+} complexes correspond to effective radii of 5.9 Å and 5.7 Å, respectively. The failure of the hydrodynamic model may not apply to real CaEDTA^{2-} , however, since tightening up the structure would result in longer correlation times. Also, with the internal motions much more restricted the correlation times of any intramolecular vector, such as the methylene C-H vectors,

would more or less equal those of the overall reorientation, cf. Table VII.

The motion of the CaEDTA^{2-} complex as a whole is determined by the solvent, i. e. by the dynamic properties of water itself. Comparison of experimental and simulation data on water dynamics can be used to correct the simulation results on the dynamic properties of EDTA. The simulation overall rotational correlation time, τ_2 , for CaEDTA^{2-} multiplied by the ratio between experimental and simulation values for the water rotational correlation time or diffusion coefficient gives ≈ 39 ps. On assumption of the hydrodynamic model, we may instead multiply by the ratio between experimental and simulation water viscosities, to give a result of ≈ 52 ps. These considerations in conjunction with the assumption of restricted internal mobility would also account for the discrepancy between the simulation and NMR values of the acetic acid C-H vector correlation times.

Conclusion

From the data presented in this work we conclude that quasiergodicity is reached relatively quickly, and that the used equilibration times are adequate. For bulk water and the solvated calcium ion, static properties seem to be plausibly represented by the present potential, as are also qualitatively dynamic properties. The failure to quantitatively reproduce dynamic, experimental data is mainly due to the absence of dynamic data in constructing the intermolecular potential. A water model that allows for intramolecular motion while correctly reproducing dynamic properties is desirable but lacking. Also for the larger EDTA molecule, structural properties are qualitatively reproduced, but are more sensitive to the imperfections of the potential. The binding of the calcium ion to EDTA is adequately treated in most respects. The CaEDTA^{2-} overall mobility is too large, but this may be corrected for by use of the known properties of the water model and experimental data. While shortcomings of the intramolecular EDTA potential render the simulation dynamics locally incorrect, the mechanism of this seems well understood and can be taken into account in the interpretation of simulations. We think the present simulations contribute substantially to the understanding of calcium binding, which, in view of its biological importance, certainly merits further investigation.

Acknowledgement is due to Bo Jönsson for valuable discussion, and to The Swedish Natural Science Research Council for financial support (grant no. K-KU 4356-104).

Registry No. H_4EDTA = N,N'-1,2-ethanediylbis[N-(carboxymethyl)]- glycine [60-00-4].

References

- (1) Levine, B. A. & Williams, R. J. P., in "Calcium and Cell Function" (Cheung, W. W., ed.); Academic Press: London, 1982; pp 1-38.
- (2) Perry, S. V., *Biochem. Soc. Trans.* **1979**, *7*, 593-617.
- (3) Klee, C. B., Crouch, T. H. & Richman, P. G., *Ann. Rev. Biochem.* **1980**, *49*, 489-515.
- (4) Teleman, O., Ahlström, P. & Jönsson, B., *Biophys. J.* **1985**, *47*, 399a.
- (5) Teleman, O. & Jönsson, B., *J. Comp. Chem.* **1986**, in press.
- (6) Friedman, H. L., *Chemica Scripta* **1985**, *25*, 42-48.
- (7) Probst, M. M., Radnai, T., Heinzinger, K., Bopp, P. & Rode, B. M., *J. Phys. Chem.* **1985**, *89*, 753-759.
- (8) Barnett, B. L. & Uchtman, V. A., *Inorg. Chem.* **1979**, *18*, 2674-2678.
- (9) Passer, E., White, J. G. & Cheng, K. L., *Inorg. Chim. Acta* **1977**, *24*, 13-23.
- (10) Mateo, P. L., Hurtado, G. G. & Vidal-Abarca, J. B., *J. Phys. Chem.* **1977**, *81*, 2032-2034.
- (11) Drakenberg, T., *Acta. Chem. Scand.* **1982**, *A36*, 79-82.
- (12) Gear, C. W.: "Numerical Initial Value Problems in Ordinary Differential Equations"; Prentice Hall: Englewood Cliffs, N. J., 1971; pp 148-154.
- (13) Hansen, J. P. & MacDonald, I. R.: "Theory of Simple Liquids"; Academic Press: London, 1976; pp 19-21.
- (14) Lennard-Jones, J. E., *Proc. Roy. Soc. London A* **1924**, *106*, 463-477.
- (15) McCammon, J. A., Wolynes, P. G. & Karplus, M., *Biochemistry* **1979**, *18*, 927-942.
- (16) Hermans, J., Berendsen, H. J. C., v. Gunsteren, W. F. & Postma, J. P. M., *Biopolymers* **1984**, *23*, 1513-1518.
- (17) Slater, J. C. & Kirkwood, J. G., *Phys. Rev.* **1931**, *37*, 682.
- (18) Margenau, H. & Kestner, N. R.: "Theory of Intermolecular Forces"; Pergamon Press: New York, 1969.
- (19) v. Gunsteren, W. F. & Karplus, M., *Macromolecules* **1982**, *15*, 1528-1544.
- (20) Berendsen, H. J. C., Postma, J. P. M., v. Gunsteren, W. F. & Hermans, J., in "Intermolecular Forces" (Pullman, B., ed.); D. Reidel: Dordrecht, 1981; pp 331-342.
- (21) Dolphin, D. & Wick, A. E.: "Tabulation of Infrared Spectral Data"; Wiley Interscience: New York, 1977.
- (22) Herzberg, G.: "Molecular Spectra and Molecular Structure: Infrared and Raman Spectra of Polyatomic Molecules"; van Nostrand: Princeton, 1945.
- (23) Reimers, J. R. & Watts, R. O., *Chem. Phys.* **1984**, *85*, 83-112.
- (24) Reimers, J. R. & Watts, R. O., *Chem. Phys.* **1984**, *91*, 201-223.
- (25) Postma, J. P. M.: "A Molecular Dynamics Study of Water"; diss., Univ. of Groningen, 1984.
- (26) John, I. G., Bacskaý, G. B. & Hush, N. S., *Chem. Phys.* **1980**, *51*, 49-60.
- (27) Matsuoka, O., Clementi, E. & Yoshimine, M., *J. Chem. Phys.* **1976**, *64*, 1351-1361.
- (28) Impey, R. W., Klein, M. L. & McDonald, I. R., *J. Chem. Phys.* **1981**, *74*, 647-652.
- (29) Impey, R. W., Madden, P. A. & McDonald, I. R., *Mol. Phys.* **1982**, *46*, 513-539.
- (30) Narten, A. H., *J. Chem. Phys.* **1972**, *56*, 5681-5687; Narten, A. H., Danford, M. D. & Levy, H. A., *Disc. Far. Soc.* **1967**, *43*, 97-107; Narten, A. H. & Levy, H. A., *J. Chem. Phys.* **1971**, *55*, 2263-2269.
- (31) Krynicki, K., Green, C. D. & Sawyer, D. W., *Discuss. Farad. Soc.* **1978**, *66*, 199-208.
- (32) Postma, J. P. M., Berendsen, H. J. C. & Haak, J. R., *Farad. Symp.* **1983**, *17/9*, 1A-7A.
- (33) Jonas, J., DeFries, T. & Wilbur, D. J., *J. Chem. Phys.* **1976**, *65*, 582-588.
- (34) Marshall, A. G.: "Biophysical Chemistry; Principles, Techniques and Applications"; John Wiley & Sons: New York, 1978, pp 712-715.
- (35) Bopp, P., Jancso, G. & Heinzinger, K., *Chem. Phys. Lett.* **1983**, *98*, 129-133.
- (36) Jancso, G. & Bopp, P., *Z. Naturforsch.* **1983**, *A*, 38A, 206-213.
- (37) Hewish, N. A., Neilson, G. N. & Enderby, J. E., *Nature* **1982**, *297*, 138-139.
- (38) Neilson, G. W. & Enderby, J. E., in "Water, A Comprehensive Treatise", (Franks, F., ed.), Vol. 6; Plenum Press: New York, 1979, pp 1-46.
- (39) Impey, R. W., Madden, P. & McDonald, I. R., *J. Phys. Chem.* **1983**, *87*, 5071-5083.
- (40) Horrocks, W. DeW. & Sudnick, D. R., *Acc. Chem. Res.* **1981**, *14*, 384-392.
- (41) See for instance C. R. Cantor & P. R. Schimmel: "Biophysical Chemistry, Part II: Techniques

for the Study of Biological Structure and Function"; Freeman: San Francisco, 1980; pp 763-788. The procedure followed here corresponds to eq. (13-108) of that ref., but in real space and with $\kappa=1$ and W_{hkl} equal to the relative masses of the atoms.

(42) NMR data, such as those in Ref. 11, yield $k_{\text{off}} < 10 \text{ s}^{-1}$.

(43) Harada, S., Funaki, Y. & Yasunaga, T., J. Am. Chem. Soc. **1980**, 102, 136-139.

Tables

Table I. σ_{ij} parameters of the Lennard-Jones potential (Eq. 2). Values are given in Å. The numbers 1-8 refer to the atom types, as given in Fig. 1.

	1	2	3	4	5	6	7	8
1	3.17	0.0	3.36	3.01	3.19	3.28	2.48	3.05
2		0.0	0.0	0.0	0.0	0.0	0.0	0.0
3			3.55	3.20	3.38	3.46	2.66	3.24
4				2.85	3.03	3.12	2.32	2.90
5					3.21	3.30	2.49	3.07
6						3.39	2.58	3.16
7							1.78	2.36
8								2.94

Table II. ϵ_{ij} parameters of the Lennard-Jones potential (Eq. 2). Values are given in kJ·mol⁻¹. Numbers 1-8 refer to the atom types, as given in Fig. 1.

	1	2	3	4	5	6	7	8
1	0.638	0.0	0.060	1.28	0.719	0.703	0.873	0.801
2		0.0	0.0	0.0	0.0	0.0	0.0	0.0
3			0.009	0.107	0.061	0.063	0.065	0.071
4				2.70	1.48	1.42	2.00	1.63
5					0.824	0.801	1.01	0.897
6						0.783	0.933	0.875
7							2.09	1.17
8								0.998

Table III. Covalent potential parameters (Eq. 3). Values of r_e are given in Å, those of B in $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$, those of α_e in degrees, those of A in $\text{J}\cdot\text{mol}^{-1}\cdot\text{deg}^{-2}$ and those of C_{1-3} in $\text{kJ}\cdot\text{mol}^{-1}$. Numbers 1-8 refer to the atom types, as defined by Fig. 1.

Bonds	r_e	B
1-2	1.00	2318.
4-5	1.22	1883.
5-6	1.52	1674.
6-7	1.03	1912.
6-8	1.46	1883.

Bond angles	α_e	A
2-1-2	109.47	58.33
4-5-4	129.0	63.73
4-5-6	115.9	50.99
5-6-7	109.5	60.92
5-6-8	110.5	114.72
6-6-7	109.5	60.92
6-6-8	109.6	63.73
7-6-7	109.5	60.92
7-6-8	109.5	60.92
6-8-6	108.9	89.22

Dihedral angles	C_1	C_2	C_3
4-5-6-7	0.0	0.0	0.0
4-5-6-8	0.0	0.0	0.42
7-6-6-7	0.0	0.0	0.0
7-6-6-8	0.0	0.0	0.0
8-6-6-8	0.0	0.0	2.09
5-6-8-6	0.0	0.0	-1.26
7-6-8-6	0.0	0.0	0.0

Table IV. Simulation parameters and thermodynamic averages. Energies are given in MJ per mol of systems. "Potential energy" refers to the total potential energy and is the sum of "noncovalent", "bond", "bond angle" and "dihedral angle" energies, cf. Eqs. 2 and 3. *At the time of the "H" simulation a new compiler had been installed, cutting computing time by some 13%.

	Water	Calcium	"C"	"M"	"H"
Length of box side (Å)	18.6	15.5	21.7	21.7	21.7
Potential cut-off distance (Å)	8.5	7.5	10.0	10.0	10.0
Equilibration time (ps)	31.3	24.6	16.93	18.56	7.68
Length of trajectory (ps)	57.6	40.0	46.08	46.08	46.08
Sampling interval (ps)	0.096	0.100	0.096	0.096	0.096
No. of neighbour pairs in thousands	83	32.6	194	197	195
Cray 1A cpu time (h)	3.01	0.61	5.10	5.18	4.42*
Cpu per pair interaction (μs)	2.72	3.37	2.47	2.46	2.12*
Temperature (K)	301±1	311±2	303±1	303±1	302±1
Pressure (kbar)	1.71±0.2	2.02±0.18	1.68±0.33	1.66±0.37	1.55±0.22
Potential energy (MJ/mol)	-8.67±0.05	-6.18±0.03	-15.29±0.07	-15.49±0.07	-12.89±0.07
Non covalent energy (")	-9.78±0.06	-6.72±0.04	-16.87±0.08	-17.10±0.08	-14.54±0.07
Bond energy (")	0.51±0.01	0.20±0.01	0.60±0.01	0.61±0.01	0.68±0.01
Bond angle energy (")	0.60±0.01	0.34±0.01	0.97±0.03	0.98±0.03	0.95±0.02
Dihedral angle energy (")	0.	0.	0.01±0.003	0.01±0.001	0.01±0.002

Table V. Correlation times, τ_2 , for water. All values in ps. The values for "long times" have, in all cases, been obtained by fitting an exponential from 0.5 to 2.0 ps. "Integral" values have been corrected for the initial rapidly decaying component and correspond to the integral of the entire tcf. The values for rigid SPC are from Ref. 25, but the integral correction was applied by us and is the same as that for flexible SPC water. MCY values are from Ref. 28, and the experimental ones from Ref. 33. The simulation temperature was 301.3 K, and all other values have been exponentially interpolated to that temperature.

		This work	Rigid SPC	MCY	Experimental
$\vec{n}_1$,	long times	0.79	1.5	1.36	
	integral	0.56	1.06	0.78	
$\vec{n}_2$,	long times	0.89	1.8	2.04	
	integral	0.63	1.27	1.48	2.06
$\vec{n}_3$,	long times	0.63	---	1.34	
	integral	0.44	---	0.76	

Table VI. Comparison of configurations by means of R factors (Eq. 8). All R factors are given in Å. The designations XC, XM, C_{bef} , C_{aft} , M, and H refer to sets of configurations and are defined in text. For R(a,b) values where a or b refers to a trajectory, the given values have been averaged over equidistant points, with a spacing of 0.96 ps, from the trajectory concerned. Where both a and b refer to trajectories, R has been averaged over both, but with a spacing between points of 3.84 ps.

	XM	C_{bef}	C_{aft}	M	H
XC	1.36	1.1±0.3	2.1±0.2	1.2±0.3	2.0±0.2
XM		0.9±0.2	1.6±0.2	0.8±0.2	1.5±0.2
C_{bef}				0.9±0.2	1.4±0.4
C_{aft}				1.6±0.3	1.3±0.3
M					1.4±0.3

Table VII. Dynamic properties of EDTA. Correlation times are given in ps, diffusion coefficients in $10^{-9} \text{ m}^2/\text{s}$. The "overall" values were obtained from the sum of the correlation functions of the EDTA Ca-N', the Ca-N'' and the N'-N'' vectors. The uncertainty for the diffusion coefficients is given as the range of values obtained for different parts of the trajectory. The experimental correlation time is taken from Ref. 11, the experimental diffusion coefficient from Ref. 10. Experimental data apply to CaEDTA²⁻.

		"C"	"M"	"H"	experiment
C-H,	τ_1	26. $\pm$ 5. 15.-38.	28. $\pm$ 4. 19.-41.	14.	
	τ_2	7. $\pm$ 1. 5.-10.	8. $\pm$ 1. 7.-11.	4.	58.
N-N,	τ_1	42. $\pm$ 3.	45. $\pm$ 6.	15.	
	τ_2	13. $\pm$ 1.	14. $\pm$ 2.	6.	
"Overall",	τ_1	49. $\pm$ 4.	47. $\pm$ 9.		
	τ_2	15. $\pm$ 1.	15. $\pm$ 2.		
Diffusion coeff.		1.0 (0.8-1.6)	1.1 (0.4-2.4)	2.	0.59

Figure captions

Fig. 1. The EDTA⁴⁻ molecule, the Ca²⁺ ion and one water molecule. Figures designate the atom types. All atoms of the same type are assumed to have identical interaction parameters, see Tables I-III.

Fig. 2. Radial distribution function for the oxygen-oxygen distances, obtained from the simulation of pure water.

Fig. 3. Unit vectors of a perpendicular, molecular coordinate system for water.

Fig. 4. Second order time correlation functions, $C_2(\Delta)$, for molecular vectors of water, as defined in Fig. 3. Solid line refers to $\hat{n}_1$, dot-dashed to $\hat{n}_2$ and dotted to $\hat{n}_3$.

Fig. 5. Radial distribution function for calcium-oxygen distances, obtained from the simulation of Ca²⁺ in water.

Fig. 6. Radial distribution functions for the distances between the calcium ion and three different sets of atoms, obtained from the "M" simulation. The first set contains the eight EDTA oxygen atoms (solid line), the second set contains the two EDTA nitrogen atoms (dashed line) and the third set contains all water oxygens (dotted line).

Fig. 7. Radial distribution function obtained as the sum of the three radial distribution functions of Fig. 6. The solid line represents the function obtained from the "C" simulation and the dashed line that obtained from the "M" trajectory.

Fig. 8. Second order time correlation functions, $C_2(\Delta)$, obtained from the "M" trajectory. $C_2(\Delta)$ for the EDTA nitrogen-nitrogen vector (solid line) and for the eight EDTA acetic acid C-H vectors (dashed line) are shown.

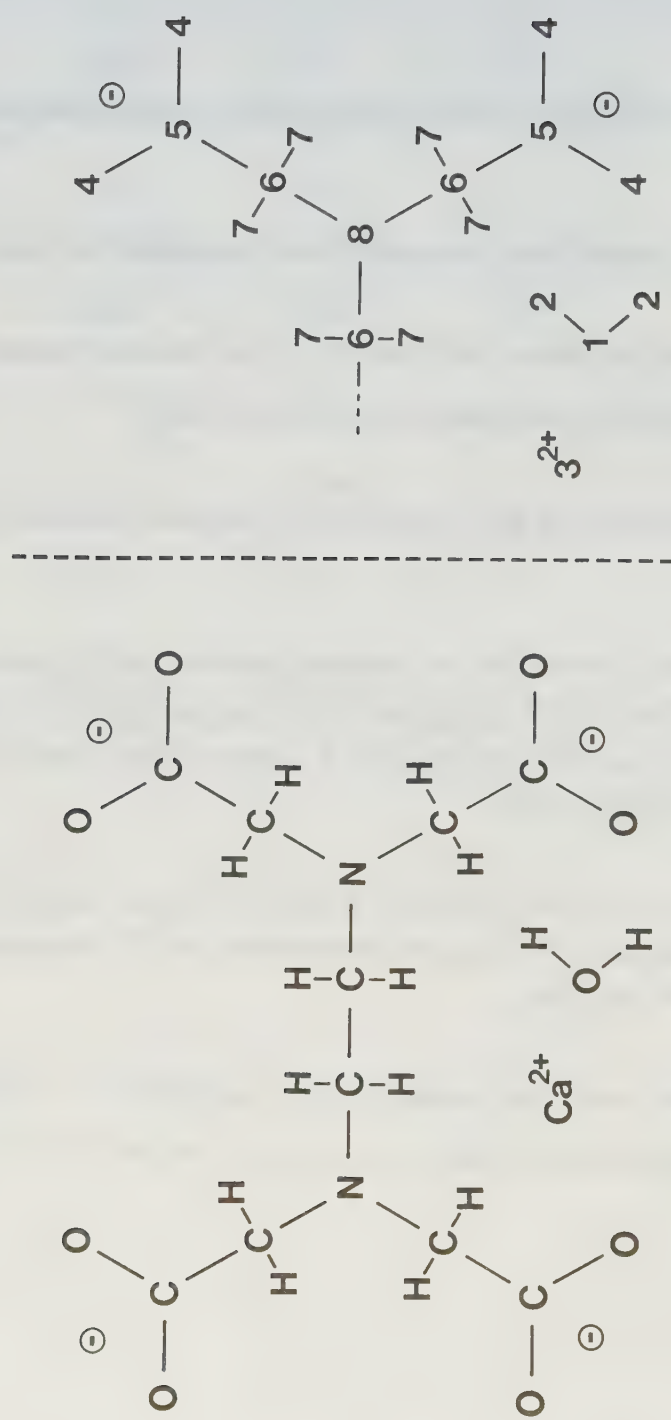


Fig. 1.

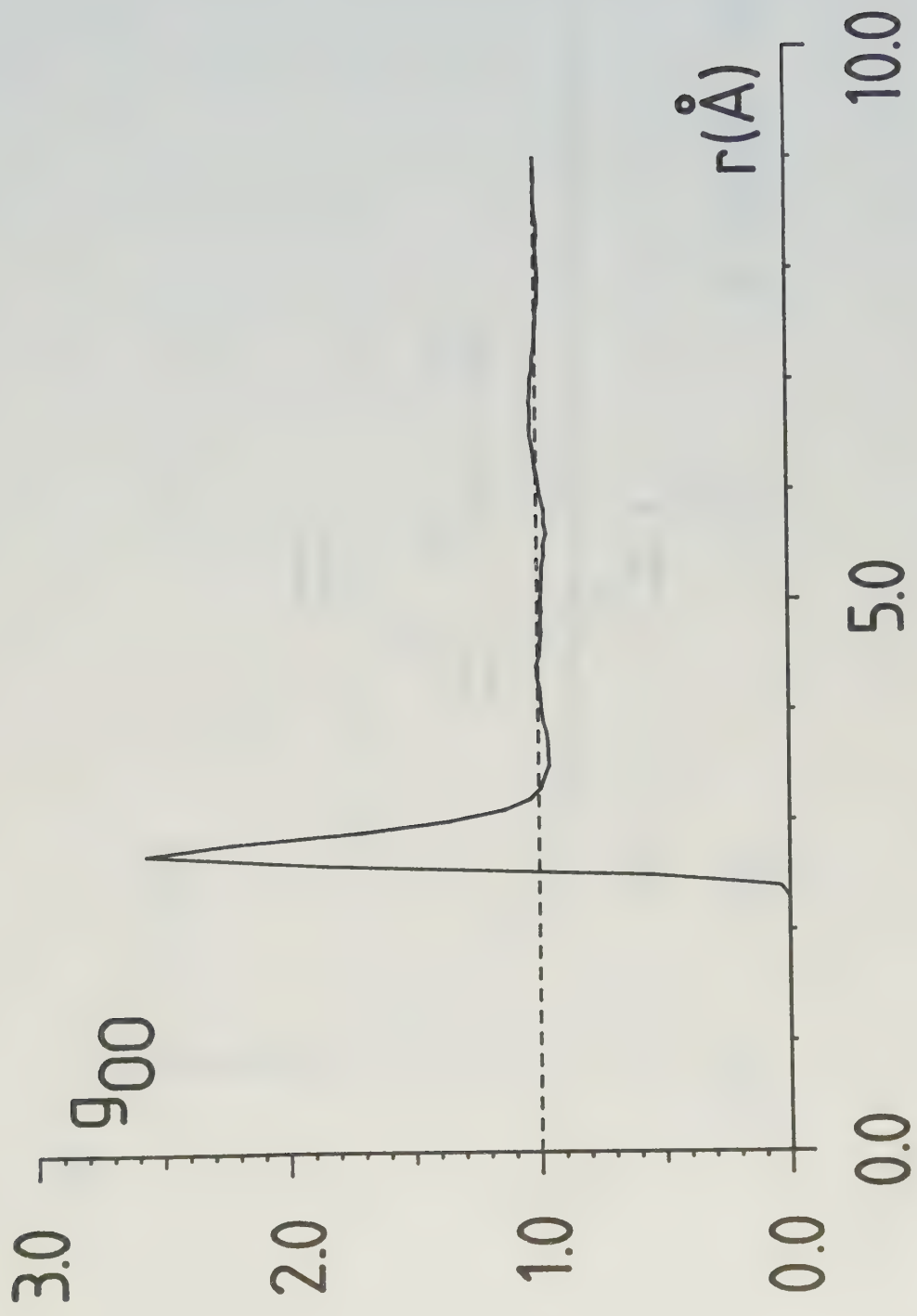


Fig. 2.

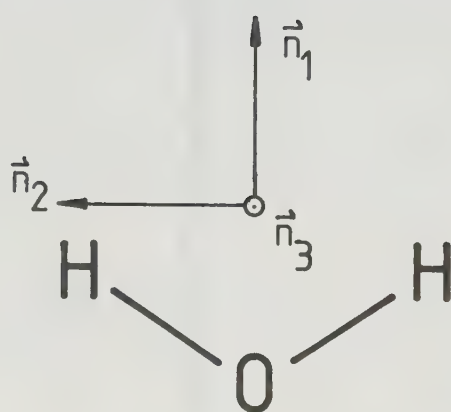


Fig. 3.

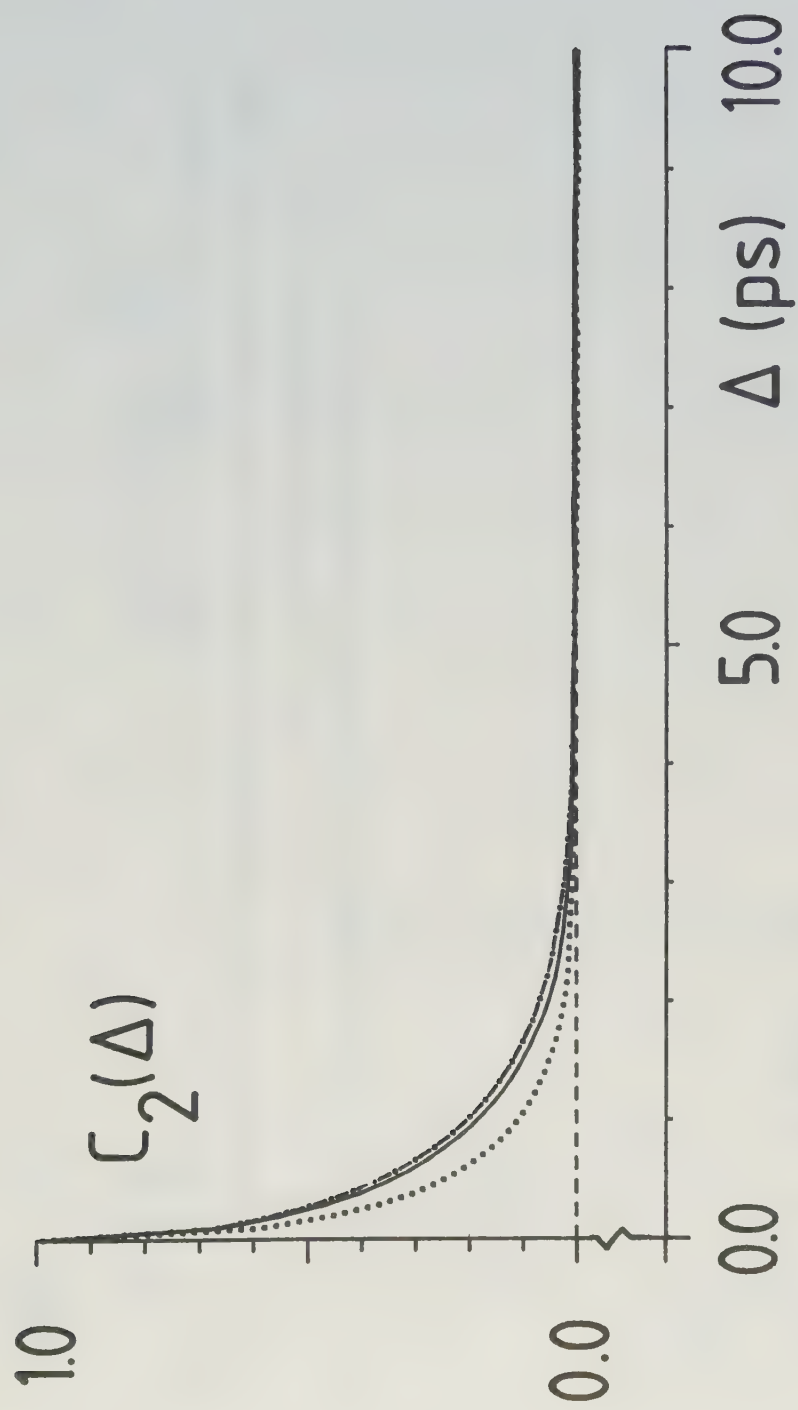


Fig. 4

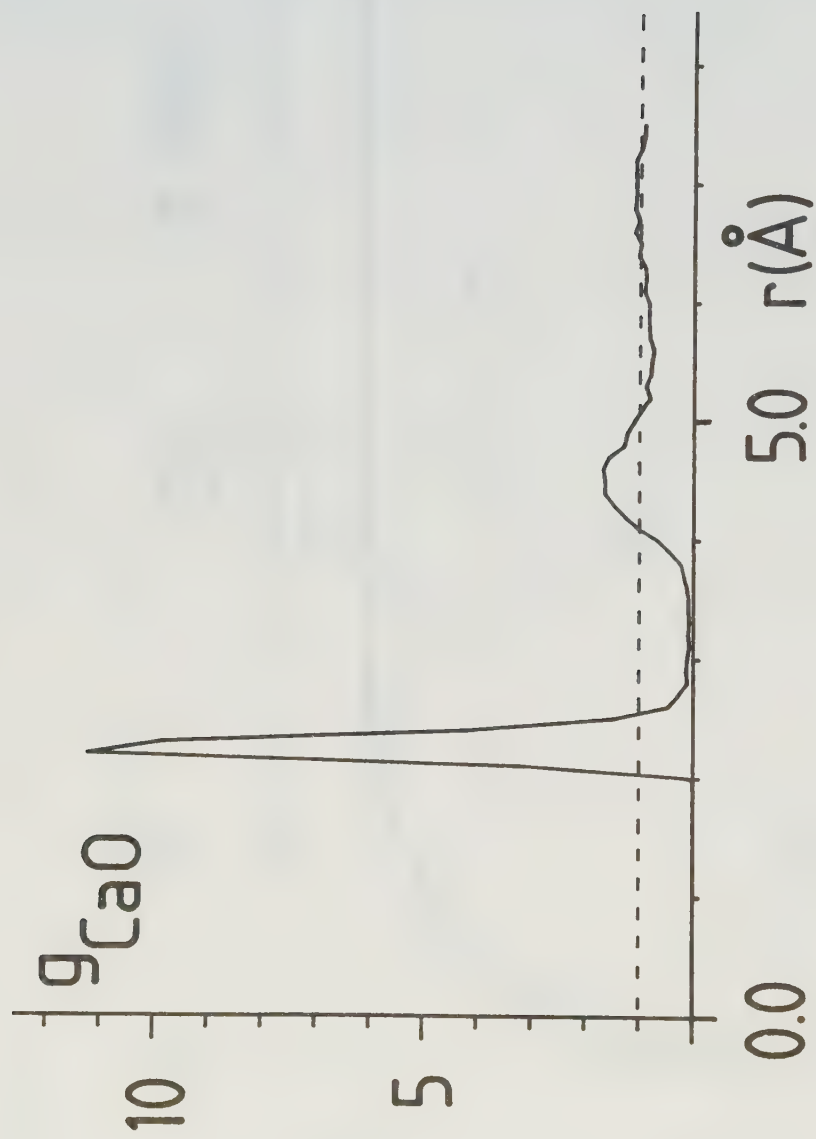


Fig. 5.

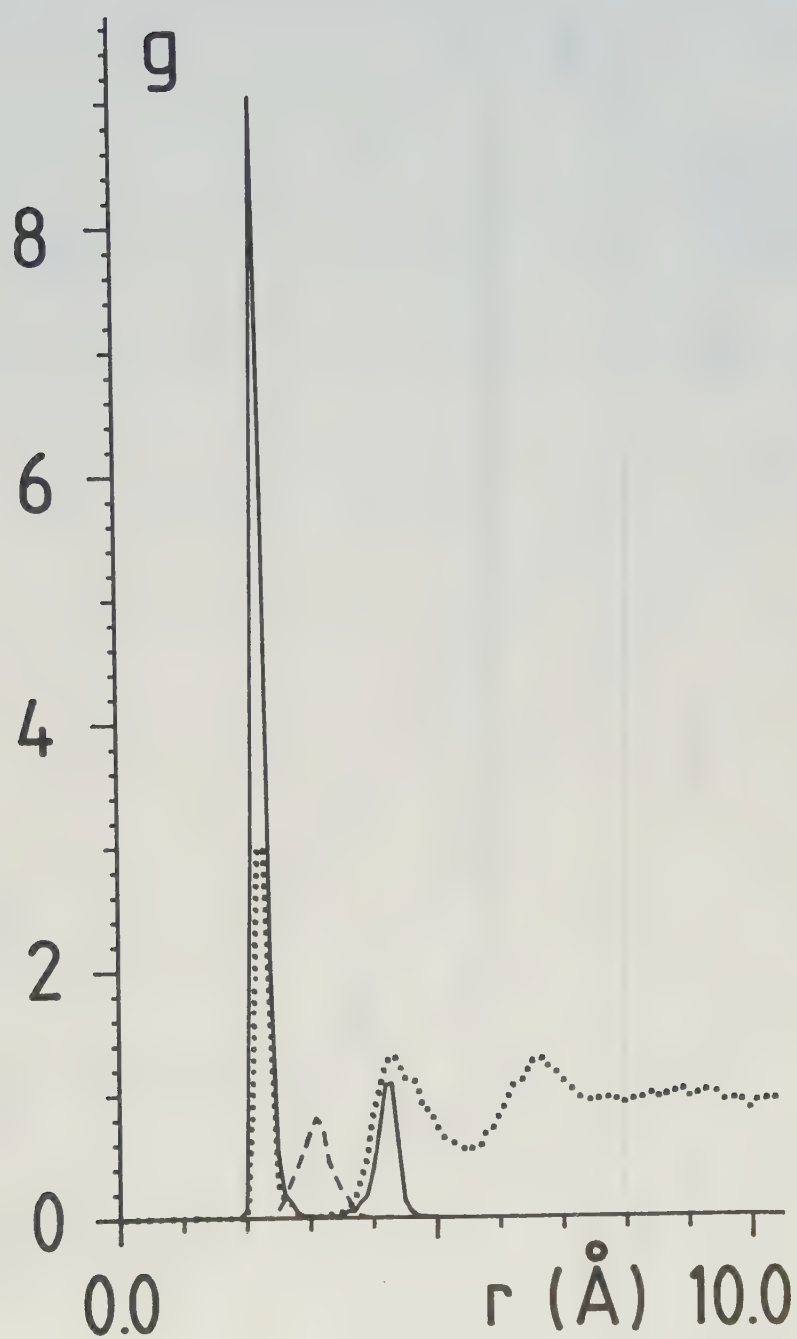


Fig. 6.

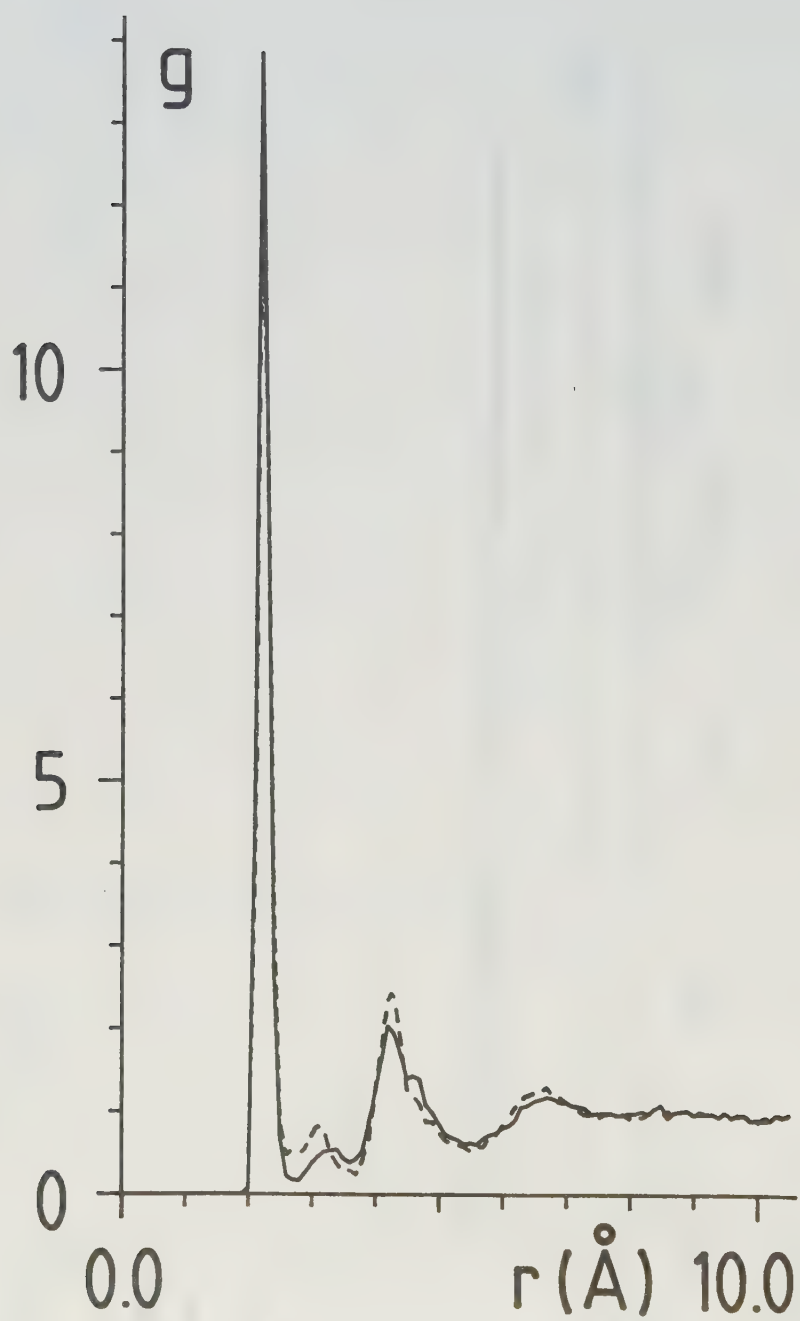


Fig. 7.

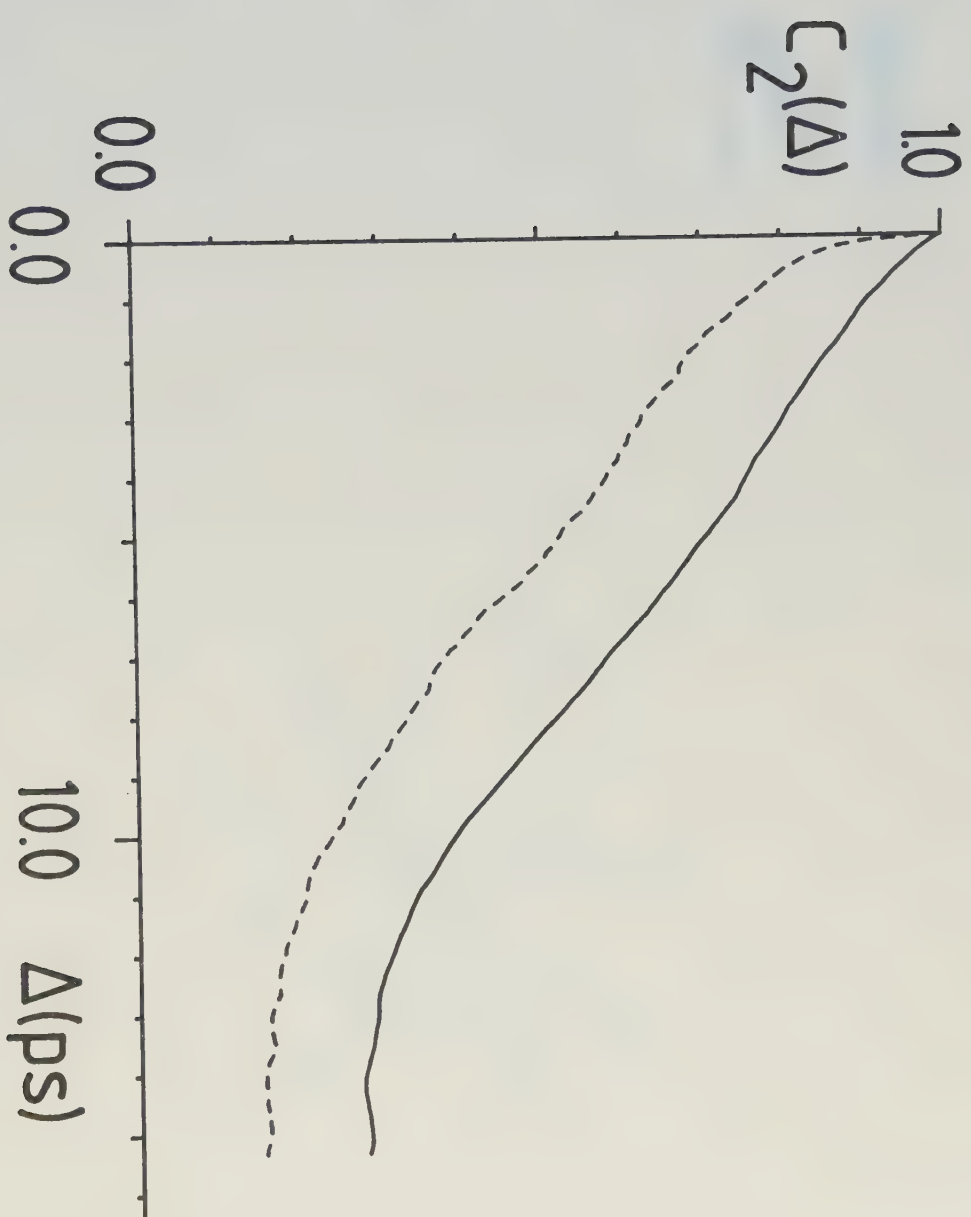
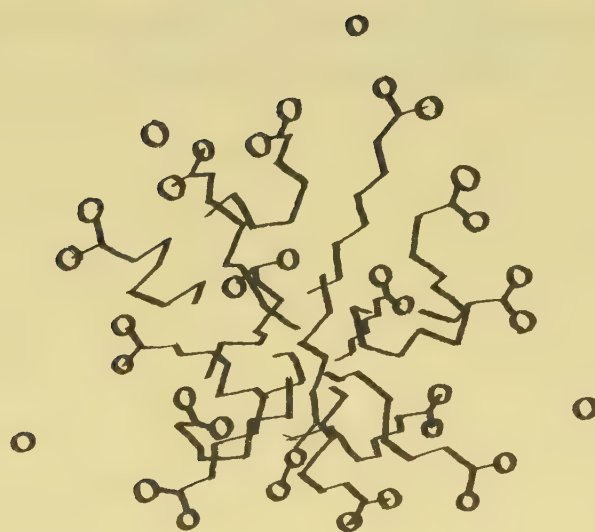


Fig. 8.

M



Molecular Dynamics Simulations of a Sodium Octanoate Micelle in Aqueous Solution

Bo Jönsson^{*}, Olle Edholm[#] and Olle Teleman^{*}

**^{*}) Dept. of Physical Chemistry 2, Chemical Centre, POB 124
S-22100 Lund, SWEDEN**

**[#]) Dept. of Theoretical Physics, Royal Institute of Technology
S-100 44 Stockholm, SWEDEN**

Abstract

Molecular Dynamics simulations of a sodium octanoate micelle in aqueous solution are reported. Simple analytical expressions have been used for the interatomic interactions, including a Lennard-Jones term and a Coulombic interaction between partial charges on each site. Two different interaction potentials have been investigated, and amphiphilic aggregation is shown to depend strongly on the electrostatic properties of the water model. For both potentials the micelle shows a broad transition region between the aqueous and hydrocarbon regions. The hydration of different carbon atoms in the chain is largest for the carboxylic atom, then decreases along the chain, reaching a minimum and then increases again at the end of the chain. The micellar translational and rotational diffusion are too fast, probably due to deficiencies in the water model. The rotational diffusion of the entire micelle is found to be an order of magnitude slower than that of the monomers. Reorientational time correlation functions and order parameters for C-H vectors in the chains are reported and agree qualitatively with a recently proposed two-step model for NMR relaxation in micelles. The dynamics of water molecules close to the micellar surface is found to be similar to bulk water.

Introduction

Molecules that contain both a hydrophilic group, ionic or polar, and a hydrophobic part, for example, a hydrocarbon tail, are termed amphiphiles. When dissolved in water, amphiphilic molecules display a number of interesting features with far-reaching consequences. Above a certain concentration, the "critical micelle concentration" (cmc), they self-associate and form clusters and large aggregates. With increasing amphiphile concentration the aggregates tend to grow and form cylindrical rods and or planar semi-infinite assemblies in an ordered structure. The micelle, originally named by McBain¹ in 1913, refers to the spherical or ellipsoidal type of assembly first formed above cmc. Its detailed structure has been subject to controversy²⁻⁹, and according to Butcher and Lamb¹⁰, "...micelle research has been known for its outlandish hypotheses". Here, we follow a more down-to-earth approach, and report a Molecular Dynamics (MD) simulation of a small aggregate consisting of 15 amphiphiles with surrounding water and counterions.

Micellar aggregates have been probed by a number of experimental methods, such as NMR^{11,12}, small angle neutron scattering¹³, diffusion measurements¹⁴ and others¹⁵. None of these methods provide enough information to deduce a complete structure or even provide conclusive evidence about a proposed structure. However, they can rule out certain suggestions and they give direct evidence about certain structural and dynamic properties of the system. Theoretical methods have the advantage that they can provide a lot more detailed information than experiment. Since these systems are complicated, however, any successful theoretical treatment involves simplifications that are not immediately justifiable. It is therefore important to compare with experimental results and adjust the model if it fails to reproduce them.

There has recently been a number of articles that investigate micelles by statistical mechanical methods¹⁶⁻²⁴. The main idea behind these methods is that the chains are packed to give a uniform hydrocarbon density within a spherical micelle, with the charged (or zwitterionic) head groups arranged on a spherical shell. An appropriate lattice is introduced, and the problem of packing hydrocarbon chains on this lattice can be solved, after some simplifying assumptions, with a moderate amount of computation. However, it is difficult to assess the validity of the assumptions introduced. Common to these statistical mechanical treatments, is the somewhat crude approximations used for the solvent and the electrostatic interactions between ionic species.

Another statistical mechanical approach to the structure and dynamics of a micellar solution is to apply a dielectric continuum model, approximating the solvent and hydrocarbon chains by the appropriate dielectric permittivity. This technique was pioneered by Overbeek and Stigter²⁵, who used the Poisson-Boltzmann approximation to find the electrostatic free energy. Later contributions along these lines have been presented by Tanford²⁶, Gunnarsson et al.²⁷ and Evans et al.²⁸. Recently there has been renewed interest in electrostatic interactions, and the problem of electrostatic interactions in micellar systems has been approached by means of powerful modern numerical techniques^{29,30}. One shortcoming of this approach is the use of a continuum approximation and

concomitant assumptions. More detailed models can be treated in MD simulations at the price of cumbersome numerical calculation. One should also realize, however, that an MD simulation involves a fair amount of modelling. This applies especially to the choice of potentials for the interactions. We will see later, that the system treated in this study is sensitive to the choice of potentials and that it is not immediately apparent how this choice should be made.

It is only within the last few years, that it has been possible to study systems with MD methods of the present degree of complication and size. Such investigations have already contributed to the understanding of planar bilayers^{31,32}. These simulations were considerably smaller than those reported in this work, since they were performed without water. Instead, the bilayer shape was maintained by means of a force which kept the head groups of the lipid molecules confined to an "average plane". This way of constraining the structure worked for planar bilayers, and provided interesting information about the organization of chains in their interior. However, it could not give information about the water hydrocarbon (head group) interface. In the micellar case this is even more critical as the interface is relatively larger for a spherical aggregate, and probably also more important for the interior structure of the micelle. Studies of micelles in the absence of water have been done with MD^{23,24} and Monte Carlo^{16,17} methods. In these, it was necessary to introduce some *ad hoc* assumption about the structure of the surface, e.g., that head groups lie on a spherical shell or that there is a certain mean square spread of head groups around a spherical surface. Such assumptions will probably affect the structure more strongly in the micellar case than in the planar. We have therefore included water and counterions in our simulation, although this increases the computing time by an order of magnitude.

Interaction Model

The inter- and intramolecular interactions in a micellar solution can in principle be obtained from quantum chemical calculations. Unfortunately, such a straightforward approach is computationally too expensive today, and probably in the near future. If one only considers the interaction between a pair of molecules, such calculations may be feasible, but still very computer demanding. The resulting intermolecular potential has then to be represented with some functional form in order to be applicable in subsequent statistical mechanical calculations. This is in principle a rather trivial fitting problem, but in practice severely hampered by the fact that the intermolecular potential has no *a priori* functional form, as it is not obtained analytically. One is therefore forced to assume a particular form with variable parameters to be determined in a least-squares fit. In typical cases, the deviation between the fitted potential and the quantum chemical calculations will amount to several kJ/mol. Some of these technical problems can be partly overcome for small molecules. For large molecules, however, and in particular molecules with internal degrees of freedom, it is extremely difficult to obtain accurate analytical representations of the intermolecular potential

surface. In order to avoid these difficulties we have chosen a semi-empirical approach, in which the fine details of the interactions have been neglected, but where we have tried to retain the essential physical properties of the intermolecular potential. Although this may seem less attractive from a fundamental theoretical point of view, we believe that it gives as good, or even better potentials than do extensive quantum chemical calculations, for systems of this degree of complexity.

In general, formation of micelles from amphiphiles is governed by a delicate balance between two competing effects: a free energy term associated with the dissolution of the hydrocarbon tails in water, favouring aggregation, which is counteracted by the head group to head group repulsion. With these effects in mind we have devised an intermolecular potential based on site-site interactions, where a site can be an atom or a group of atoms (a pseudoatom). The only pseudoatoms in the model are the amphiphilic $-\text{CH}_2-$ and $-\text{CH}_3$ groups. A pair of sites i and j belonging to different molecules or two sites within a molecule more than three bonds apart will interact via a Lennard-Jones (LJ) plus a Coulomb term

$$U_{ij}(r) = 4\epsilon_{ij} [(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^6] + q_i q_j / (4\pi\epsilon_r\epsilon_0 r_{ij}) \quad (1)$$

The ϵ_{ij} and σ_{ij} are the usual LJ parameters, given in Table I, and q_i are the partial charges. The parameters have been taken from the literature³³⁻³⁵. They are probably not the most accurate set available, but should nevertheless give a reasonable description of short ranged interactions in the system. The relative permittivity ϵ_r has been set equal to 1.

There exist several water models, which all reproduce simple thermodynamic properties such as pair correlation functions, pressure and internal energy with reasonable accuracy. In the present work we have used an empirical Simple Point Charge (SPC) model of Berendsen et al.³⁶ The reason for this choice is partly simplicity, as the SPC model conforms to the interaction model, Eq. (1). Unfortunately, there has not to our knowledge been any determination of the dielectric permittivity for SPC water. Recently one such simulation for a non-empirical water potential due to Matsuoka et al.³⁸ was presented, and its dielectric permittivity was found to be ≈ 35 at room temperature³⁷, roughly half the experimental value for water. One reason for the too low value is the neglect of nuclear and electronic polarisation. One could argue that the SPC model should behave better, since it is an effective pair potential including many-body polarisation terms in an average way. However, we do not think this is the case, as the Matsuoka potential, although non-empirical, has been fitted to a site-site model with a dipole moment of ≈ 2.2 D, considerably larger than the experimental gas phase value, and close to the dipole moment of an SPC water molecule. Simulations of a polarisable Stockmayer fluid have shown that the dielectric permittivity

is doubled on inclusion of electronic polarisation³⁹. Thus we do expect the SPC water to inadequately screen ionic interactions in the system.

From the above considerations we may forecast that a straightforward application of the potential model as given in Tables I and II (in the following referred to as the Full Charge (FC) model) will not lead to a proper aggregation of amphiphiles. One very simple way to remedy the lack of dielectric screening in the water potential is to scale ionic interactions with the high frequency dielectric constant $\epsilon(\infty)$, which is of the order of 2-5. To test this approach, a second set of simulations was performed with the Na^+ and -COO^- charges reduced to half their real values, that is +0.5 and -0.5 of an electron charge. This set of simulations will be referred to as the Reduced Charge (RC) model. Some violence has been made to the molecular model by introducing this *ad hoc* screening, but, as discussed above, there are arguments in favour of such a procedure.

The use of CH_2 and CH_3 pseudoatoms means that the dipole moment associated with the C-H bonds is neglected as is also the electronic polarisation of the hydrocarbon chains. These approximations are probably less critical for the micelle formation than the choice of water model.

Apart from site-site interactions, the carbon skeleton also includes a dihedral potential. We have used a third order Fourier expansion

$$u(\varphi) = \sum_{k=1}^3 a_k \cos(k\varphi) \quad (2)$$

with a_k equal to 11.79, 9.84 and 10.60 kJ/mol for $k = 1, 2$ and 3 respectively for the C-C-C-C dihedrals (the cis-configuration corresponds to $\varphi = 0^\circ$). For the O-C-C-C dihedral angle we used $a_3 = 0.42$ kJ/mol and $a_1 = a_2 = 0$.

The internal bond and bond angle vibrations have been treated explicitly, and the harmonic force constants used are given in Table II. The numbering of atoms is explained in Fig. 1. One of the advantages of including these is to obtain the contribution from the nuclear polarisation, although only classically, to the dielectric permittivity.

Simulation Technique

The computer program used in the present study has been described in detail elsewhere⁴⁰. Here, only a short account will be given of the technique. The program is based on independent atoms or pseudoatoms and hence includes covalent interactions in terms of harmonic or periodic

potentials. This usually requires a rather small time step when integrating Newton's equations of motion, due to the rapid bond and bond angle vibrations. In order to circumvent this problem we have devised a double time step algorithm⁴⁰, where the rapidly varying degrees of freedom are integrated with a small time step, while for the slowly varying ones a larger time step can be used. In practice this means that bond and angle vibrations are integrated with $\Delta t = 0.2$ fs (fs = 10^{-15} seconds) while $\Delta T = 1.2$ fs is used for the remaining degrees of freedom. In the present case less than ten per cent of the total computation time was spent calculating the rapid motions, this is thus an efficient way to handle flexible molecules.

Both simulations (FC and RC) were started from a structure with the hydrocarbon chains in an all trans conformation and extending radially from the coordinate origin. This corresponds to a fairly loose micellar assembly. The water molecules were placed on the bases of a primitive cubic lattice in the volume outside the amphiphiles. Fifteen randomly chosen water molecules were exchanged for sodium ions.

The number of water molecules included in the simulations was 1094 which together with 15 sodium octanoate monomers add up to 3447 atoms. A cubic simulation box with a side length of 34.2 Å was used together with periodic boundary conditions. A spherical cutoff radius of 10 Å was used and in order to speed up the force evaluation a neighbour list was employed, which was updated with an interval of 4.8 fs. Due to truncation of long ranged electrostatic interactions the total energy is no longer a constant of motion. To remedy the drift in energy, velocities were scaled with an interval of 24 fs giving an average temperature of 294 K in both simulations. The analysis frequency was set as large as 100 and 120 fs, in the FC and RC case respectively, in order to reduce the amount of data to be analysed.

The two sets of simulations reported here were equilibrated for ≈ 80 ps (1 ps = 10^{-12} seconds) and the analysis of the FC simulation was 30 ps long, while the RC simulation was analysed during 72 ps. To simulate one picosecond took 19 minutes of CPU time on a Cray-1, which corresponds to 2.12 μ s per site-site interaction.

Results and Discussion

With the trajectories for both water and amphiphile molecules as well as counterions, an almost unlimited number of different analyses can be performed. It is thus necessary to limit these and one guideline has been to analyse quantities which are in principle amenable to experimental determination. We have also tried to reveal differences between the present model and others reported in the literature. It is also instructive to compare results obtained with the two different potential models, i. e., with full charges and reduced charges. This is an important point, since one aspect of the present simulations has been to explore different potential models and provide an understanding of how to find better molecular models of amphiphilic organization in aqueous

solution.

It is usually straightforward to calculate thermodynamic quantities like pressure or internal energy from simulations. There are no problems associated with their presentation either, since the averages are just a number plus some statistical error. To present structural information is more difficult and one is usually limited to a one-dimensional representation. For liquid argon this is no problem, but for water the analysis and presentation have to be simplified. For example, the many-dimensional pair correlation function is reduced to one-dimensional pair correlation functions between oxygen-oxygen, oxygen-hydrogen and hydrogen-hydrogen atoms. In the following, all pair correlation functions presented are one-dimensional, which means that they have been spherically averaged. This may obscure some of the information but mention of this will be made when appropriate.

For simplicity, this section has been divided into subsections starting with a discussion of global thermodynamic and dynamic properties of the aggregate. Following this, are two sections devoted to structural properties of the hydrocarbon chains and their dynamic behaviour, especially with respect to NMR relaxation measurements. After this we turn to a description of the solvent, and finally we analyse the counterion distribution and attempt a comparison with a continuum theory.

Aggregate structure and dynamics

Table III contains some averages of structural and dynamic properties of the aggregate. The system was kept at a total density about 10 % less than experimental, however, the pressure is still considerably higher than atmospheric. It is well known that even small inaccuracies in the short range potentials may cause large errors in the pressure. The high pressure may seem strange when knowing that rigid SPC water at normal density gives a pressure of ≈ -50 MPa³⁶. However, the introduction of intramolecular flexibility increases the pressure considerably, as has been noted in simulations of pure water⁴¹. Thus, choosing between simulation at proper pressure or proper density, we decided to compromise and reduced density somewhat. The resulting pressure is still fairly high from an experimental point of view, but we believe that it will only marginally affect most of the properties evaluated.

Apart from the pressure, some overall properties of the micelle are given in Table III. The radius of the micelle was calculated as the average distance between the centre of mass of the micelle (com) and the head group carboxyl carbons (C1 in Fig. 1). In the FC case the micelle is larger and more diffuse than in the RC case. This confirms what is more clearly seen from the snap shots of the micelle in Fig. 2. We also refer to Fig. 3 where the total hydrocarbon density is shown as a function of distance from the centre of mass. Experimental information about micelle radii and diffuseness of their structure may be obtained from neutron scattering experiments. For sodium dodecyl sulfate (SDS), Cabane et al.⁴² give an average radius of 18.4 Å with a standard deviation of 10%. In that case the micelle contained a larger number of chains, which were also longer than

in our case. The standard deviation in the RC case is comparable to theirs in absolute magnitude, but larger in relative size. The radius can also be calculated from the moment of inertia. We then calculated the radius of a solid homogeneous sphere having the same average moment of inertia as the micelle. This gives in the FC case 14.7 Å, a value far too large, while we get 10.4 Å in the RC case. The former value reflects the very loose structure of the FC micelle but also the fact that it is considerably aspherical, see Table III.

The translational diffusion coefficient, cf. Table III, is calculated from the slope of the mean square displacement against time. The values given are 5-10 times faster than experimental ones¹⁴, which are of the order of $1 \cdot 10^{-10} \text{ m}^2 \text{ s}^{-1}$. The rotational diffusion constant of a rigid spherical body can be obtained from a time correlation function (tcf) of the l :th order Legendre polynomial, $P_l(\cos\Delta\theta(\tau))$, where $\Delta\theta(\tau)$ is the change in orientation in time τ . For a diffusive motion these tcf:s decay as $\exp[-l(l+1)D_R\tau]$, which means that D_R is independent of l . We have chosen to use the second order tcf:s because the rotational diffusion of the micelle enters the NMR relaxation theory, see below, via second order tcf:s. The rotational diffusion constant can then be obtained from a linear fit to the logarithm of the tcf. A further problem is that the micelle is neither perfectly spherical nor a rigid body. In consequence it is not immediately clear how $\Delta\theta$ is best defined. As $\Delta\theta$ we have chosen to use the angular displacement of the vector from the micellar centre of mass to C1. Next, one has to consider how to separate the rotational diffusion of the micelle as an entity from that of individual monomers. We first calculate $\langle \sin^2\Delta\theta \rangle$ from the average of the squared angular displacements:

$$\langle \sin^2\Delta\theta(\tau) \rangle_t = \langle \sum_{i=1}^N [\mathbf{r}_i(t+\tau) \times \mathbf{r}_i(t)]^2 \rangle_t / N \quad (3)$$

where $\mathbf{r}_i$ are unit vectors from com to the N different C1 atoms. The brackets $\langle \dots \rangle_t$ indicate a time average. This function is sensitive both to diffusion of the entire micelle and to the diffusion of individual monomers. Eq. 3 will primarily reflect the latter since it is much faster than the former. To obtain the rotational displacement of the entire micelle, we have to use a function that only takes into account concerted rotations of all monomers. One way to achieve this is by averaging $[\mathbf{r}(t+\tau) \times \mathbf{r}(t)]$ over all chains before the square is taken, thus:

$$\langle \sin^2 \Delta\theta(\tau) \rangle_t = \langle \left[\sum_{i=1}^N \mathbf{r}_i(t+\tau) \times \mathbf{r}_i(t) \right]^2 \rangle_t / N \quad (4)$$

which describes the rotation of the whole micelle as an entity. The rotational diffusion coefficient derived from Eq. 4 is an order of magnitude smaller than the one derived from Eq. 3.

To interpret the diffusion constants one may use Stokes' formulae for a sphere of radius a in a fluid with viscosity η :

$$D_T = k_B T / (6\pi\eta a) \quad (5)$$

and

$$D_R = k_B T / (8\pi\eta a^3) \quad (6).$$

Since both diffusion coefficients are obtained from the simulation, η and a can be estimated independently. We get for the FC and RC models 15.4 Å and $0.26 \cdot 10^{-3}$ Ns/m² and 12.4 Å and $0.18 \cdot 10^{-3}$ Ns/m² respectively. It is not unreasonable that radii obtained in this way are 2-3 Å larger than those given by the average radial position of the carboxyl carbons, since the carboxyl oxygen atoms extend a few Ångströms further out into the solvent. The viscosity is 4-5 times smaller than the experimental one for water⁴³, about $1 \cdot 10^{-3}$ Ns/m², which indicates that the dynamics is too fast.

Eqs. (5) and (6) are based on a hydrodynamic model assuming "sticky" boundary conditions, i. e., that the solvent at the surface of the moving (rotating) body follows the motions of that body. In some cases it is more appropriate to assume "slip" boundaries, which means that there is no tangential force between the moving body and the solvent. There is no closed formula for "slip" boundaries and rotational diffusion, but numerical calculations for ellipsoids⁴⁴ yield diffusion constants that increase indefinitely when the ellipsoid elongation approaches zero. With an experimental value of η and a reasonable choice of a , the rotational diffusion constant could in principle be explained by introducing "slip" boundary conditions, whereas the translational diffusion could not. The latter would, in fact, still be too large by a factor of 3. Thus, the reason for the fast diffusion must be looked for elsewhere.

The moments of inertia with respect to the three principal axes are given in Table III together

with their root mean square deviation. At each time step the components of the diagonalized tensor have been ordered, largest, second largest and smallest component. The figures in Table III then represent averages over these three quantities. In the FC case the micelle is considerably aspherical. An interesting question is whether this prolate aggregate actually performs any substantial rotation or if it changes shape much faster than it rotates. We have calculated the fluctuations in the moment of inertia tensor in a laboratory frame, to obtain a hint at possible sources of the aggregate deformations. The time correlation function of the fluctuations was found to decay with a time constant of ≈ 20 ps in the RC case. This is close to the time scale for monomer diffusion. Thus we conclude that the deformations, at least initially and described in a laboratory frame, are governed by monomer diffusion rather than by rotation of the whole aggregate.

Chain conformations

We turn next to an analysis of the hydrocarbon part of the micelle. The hydrocarbon density decays smoothly from its maximum value down to zero over a distance of about 5 Å, cf. Fig. 3, which in part implies that the micelle has a rough surface, with hydrocarbon chains protruding into the water. In Fig. 3 the hydrocarbon density apparently drops near the centre of the micelle ($r < 1.2$ Å). However, statistics is poor in that region and with average hydrocarbon density, there would be 0.02, 0.13, 0.38 and 0.74 hydrocarbon groups in the $0 < r < 0.5$ Å, $0.5 < r < 1.0$ Å, $1.0 < r < 1.5$ Å and $1.5 < r < 2.0$ Å regions respectively. Thus the hydrocarbon density drop may be a statistical fluctuation. The more detailed analysis in Fig. 4 shows the density of different hydrocarbon groups as a function of distance from the centre of the micelle. For the FC case, cf. Fig. 4a, there is no significant difference between the curves, which we interpret as a result of the very loose structure of this aggregate. In the RC case there are distinct peaks for the different distributions. The peak for the C1 distribution is found furthest out and then follow C2, C3, C4, C5 and C6. Then there seems to be a reversal and we find the maxima of the C7 and C8 distributions slightly further out than the C5 maximum. Further information is given in Table IV where the average radial positions of different carbon atoms are given together with standard deviations of the distributions. This is compared to two simple models. In the first, we assume that each CH₂ group occupies a volume of 28 Å^3 and that different groups can be packed successively into spherical shells of a volume equal to the number of chains (15) times 28 Å^3 . We used the radius that encompasses half the spherical shell volume as $\langle r \rangle$ in this case. In the second model, the chains were packed in an all trans configuration. There is a qualitative difference between the simulation results and the results of these simpler models. Obviously the underlying assumptions are too crude, although the "close packed" model gives reasonable agreement for the first five carbon atom distributions.

The results can also be compared to a simulation of Haile and McConnel^{23,24}, where head groups were fixed to a spherical shell. That picture is certainly not realistic for the head group distribution and while we get a width of the radial distributions for the CH₂ groups that is roughly

constant as a function of the chain position, these authors find distributions that vary a great deal in width when going along the chain. In a second paper Woods et al.²⁴ loosened the constraint that the head groups lie on a spherical shell. They instead used a harmonic potential, that gave a spread in radial head group distribution of about $\pm 2 \text{ \AA}$ and, as a consequence, radial distributions for the different chain positions that are more similar to ours. These authors, however, report a much narrower distribution for C2, but then again a continuous increase in distribution width on going along the chain. This contradicts the data reported here, cf. Table IV, and we can, in fact, see no reason for the C2 distribution being much narrower than that for C1. Gruen¹⁹ also applied the constraint that the head groups lie on a spherical shell. He obtained distributions that successively widen on going along the chain starting at the head group. This disagrees likewise with the approximately constant distribution widths reported in this work. Similarly, both Haile & McConnel and Gruen find a monotonic decrease in the position of the maxima of the individual carbon atom distributions and not a "reversal" as found in the simulations. We believe that these differences are due to the unphysical constraints applied to the head group positions in their approaches.

Bendedouch et al.¹³ have performed small angle neutron scattering experiments that provide information on the packing of CH_3 groups in sodium dodecyl sulfate (SDS) micelles. These results have been compared to different models by Dill et al.⁷. From the experiments, a function $A(q)$ is extracted, that is the sine Fourier transform of the CH_3 density multiplied by r .

$$A(q)/A(0) = q^{-1} \int_0^{\infty} \rho(r) \sin(qr) r dr \quad (7)$$

This function is, however, only known up to a wave vector of about $0.3 \text{ \AA}^{-1}$. This means that the spatial resolution is only about $10 \text{ \AA}$. Thus it is not possible to draw any detailed conclusions on the spatial distribution from these data. We have performed the Fourier transform of the methyl densities obtained from the simulation (Fig. 5) and find that $A(q)$ is virtually identical for the FC and RC models up to about $0.3 \text{ \AA}^{-1}$. Our micelle is smaller than those of the experiments of ref. 13, leading to a curve $A(q)$ that extends to higher q values. We may, however, make a comparison if we assume that the shape of the methyl density function is the same, but rescale r with a constant a proportional to the third root of the number of methyl groups. Then we get a similar rescaling of the properly normalized Fourier transform so that $A(q)$ goes into $A(q/a)$. By rescaling to 54 chains we get, from either of our models, perfect agreement with the curve obtained by fitting the experimental values to a Gaussian in ref. 13. Further, a uniform distribution, that drops to zero at the appropriate radius, corresponds to an $A(q)$ that deviates from the experimental only outside

$0.20 \text{ \AA}^{-1}$. This illustrates that one would really need to know $A(q)$ well beyond $0.3 \text{ \AA}^{-1}$, to be able to say anything about the details of the hydrocarbon distribution.

The electron density of the system was calculated as a function of the distance from com of the micelle (cf. Fig. 6). It is less structured than what is usually assumed in models used to interpret scattering data, and it is also, at least in the hydrocarbon region, subject to statistical errors. Nonetheless, the increase in electron density at the head group is discernible, while the difference in electron density between the hydrocarbon and aqueous regions is not.

An analysis of chain conformations shows that the average fraction of trans bonds is 53% for the FC model and 50% for the RC model. Since the *gauche*⁺ and *gauche*⁻ should be equally probable we can estimate the statistical error in this figure from the different occupation numbers for two states to be around $\pm 3\%$. This means that the difference between the two models is insignificant. From the trans fraction in different bonds and their averages over all simulations (Table V), we note a small drop in trans content for the last bond in the chain. The total fraction of trans bonds corresponds to that of a model with the different bonds taking on trans and *gauche* states independently, with an excess *gauche* energy of 1.7 kJ/mol. We may also compare the number of chains with different number of *gauches*, with the simulations and then obtain Table VI, with differences that are hardly significant.

The transition rate between the different trans-*gauche* states are also estimated from the simulation, giving an average rate for each bond of 0.067 ps^{-1} ($\text{ps} = 10^{-12}$ seconds) in the FC case, and 0.038 ps^{-1} in the RC case. Since this means that each bond changes state two or three times during the whole simulation, it explains why statistics are rather poor. We have calculated the dihedral distribution $W(\varphi)$ as an average over all bonds, from which the excess dihedral free energy can be obtained, using

$$\Delta F(\varphi) = -k_B T \ln W(\varphi) - U(\varphi) \quad (8)$$

where $U(\varphi)$ is the input dihedral potential. Figure 7 shows that ΔF has maxima of about 2 kJ/mol at the *gauche* states, and that it has minima at the barriers between the states. This will decrease the fraction of *gauches* and will increase the transition rate between different states. The average root mean square end to end distance (C1-C8) of the chains has also been calculated and compared to a model with uncorrelated *gauche* states. In both the FC and RC models we find it to be about $7.15 \text{ \AA}$ which can be compared with a lengthy formula due to Flory⁴⁵ giving $6.65 \text{ \AA}$ (for a *gauche* energy of 1.7 kJ/mol), which is about 7% shorter. One would need a *gauche* energy of about 2.7 kJ/mol to reproduce the simulation result.

NMR relaxation rates

The time correlation functions of spherical harmonics describing the orientation of C-H bonds determine the NMR dipole-dipole relaxation for ^{13}C . We define

$$g_{2m}(\tau) = 4\pi \langle Y_{2m}(t + \tau) Y_{2m}(t) \rangle_t \quad (9)$$

where the spherical harmonics Y_{2m} are functions of the two polar angles that determine the direction of a C-H bond in a laboratory coordinate frame and $\langle \dots \rangle_t$ denotes a time average as well as an average over the monomers. The spin lattice NMR relaxation time T_1 , is given by the Fourier transforms $J_m(\omega)$ of these functions

$$1/T_1 = C [J_0(\omega_C - \omega_H) + 3J_1(\omega_C) + 6J_2(\omega_C + \omega_H)] \quad (10)$$

where ω_C and ω_H are the Larmor frequencies for ^{13}C and ^1H respectively⁴⁶. C is a constant containing the gyromagnetic factors (γ_C and γ_H), Planck's constant (h), the average C-H distance (r) and the permeability of vacuum, μ_0 :

$$C = [\mu_0 \gamma_C \gamma_H h / (8\pi^2 r^3)]^2 / 40\pi \quad (11).$$

The formula for the time correlation functions simplifies, since for a micelle all orientations occur with equal probability. Thus Eq. 9 can be symmetrized

$$g_2(\tau) = 0.8\pi \sum_{m=-2}^2 \langle Y_{2m}(t + \tau) Y_{2m}(t) \rangle_t$$

The use of an addition theorem for spherical harmonics now gives a formula that is independent of the choice of reference system:

$$g_2(\tau) = \langle P_2 [\cos \Delta\theta(t, t+\tau)] \rangle_t \quad (12).$$

Here $\Delta\theta(t, t+\tau)$ is the angle between the C-H vector at times t and $t+\tau$ and $P_2(\cos\Delta\theta)$ is the second order Legendre polynomial. These time correlation functions (tcf:s) start at unity and at first decay very fast, in the order of a picosecond, to about 0.6 (see Fig. 8). This is due to fast vibrational and librational motions within the same conformational state. Then follows a decay that is characterized by a time constant of the order of 2 ps. This brings the tcf:s down to about 0.05-0.1 in less than 10 ps. It is a reasonable guess that this latter part of the decay is due to trans-gauche isomerizations. However, as discussed above, the rate of trans-gauche isomerization is in the 10-20 ps regime per bond. This means that in one chain there is on average one trans-gauche isomerization each 3-5 psec. This may be a bit slow to explain the decay in the tcf:s. The decay rate at longer times is hard to obtain with good accuracy, but there seems to be other components faster than the rotational diffusion, cf. Table III.

For ^{13}C NMR relaxation in micelles, a two step relaxation model has been proposed⁴⁷. It is assumed that the function $g_2(\tau)$ in Eq. 12 is of the form

$$g_2(t) = (1 - S^2) \exp(-t/\tau_f) + S^2 \exp(-t/\tau_s) \quad (13).$$

Relaxation times, T_1 , have been obtained from subsequent ^{13}C (ref. 48) and ^2H (ref. 49) experiments performed at different Larmor frequencies, and fitted to Eq. 13. That yielded, depending on chain position and system under consideration, order parameters $S \approx 0.1-0.2$, a fast correlation time $\tau_f \approx 10-20$ ps, and a slow one $\tau_s \approx 1-2$ ns. It must, however, be stressed that from the Larmor frequencies used, only an effective average fast correlation time can be obtained. If the tcf is nonexponential for times much shorter than the inverse of the highest Larmor frequency (≈ 500 MHz), a fast correlation time cannot be determined. From the simulation we find a whole spectrum of decay times. Since the tcf:s are statistically reliable only up to about 15 ps, an average τ_f , as implied by Eq. 13, cannot be calculated. The initial decay is almost an order of magnitude faster than the τ_f reported⁴⁸. This might well be compensated for by long tails extending up to nanoseconds.

The order parameter obtained from NMR experiments on micelles⁴⁸ are calculated from a

plateau in the autocorrelation function, and obtained from measurements of the NMR relaxation time at different magnetic fields. We have tried to estimate an analogous order parameter from the simulation by averaging the tcf:s between 12 and 18 ps. Actually this corresponds to the square of the order parameter. As seen from Table VII it is roughly constant and equal to ≈ 0.2 until C7 where it drops to ≈ 0.1 . The magnitude of the order parameters agree roughly with experiments although the drop towards the end of the chain in that case is more gradual. The order parameters in ref. 48 are for sodium p-octylbenzenesulfonate, which makes the comparison less straightforward. However, Klason and Henriksson⁵² report order parameters for octanoate in a lamellar phase and they find them to be ≈ 0.2 with a sharp drop at the C7 and C8 atoms.

One can also define a "structural order parameter" from the average of the second order Legendre polynomial for the cosine of an angle between a C-H bond and a reference director. In the lamellar case it is obvious to choose the director as the normal to the lamellae. In a micelle this choice is less obvious and we have used two different directors: i) the vector between com and the hydrocarbon group under consideration, ii) the vector between com and the carboxyl carbon. These structural order parameters are smaller, see Table VII, than the ones obtained from the tcf:s and they also decrease faster along the chain. The same quantity has been calculated by Gruen¹⁹, who reports an order parameter that starts at 0.2 and then gradually decreases along the chain. Such a behaviour does not agree with our structural order parameters, but is more in line with the results obtained from the plateau of the tcf:s. In the case of Gruen's model, the order parameters for the first carbon atoms will depend on the magnitude of the "orientational weighting factor" used.

Water structure and dynamics

Water penetration into micelles has been vividly debated^{2,4,15}. It seems that most authors now agree that penetration is limited. This apparently contradicts the present results, see Fig. 9, which for both simulations show a rather broad transition region between the hydrocarbon and aqueous phases. The present octanoate micelle is rather small with only 15 monomers, in which case a slight asphericity may be favourable. As mentioned above, the asphericity obscures distribution functions of the kind shown in Fig. 9, in making them broader, and may thus partly account for the broad transition region. This is certainly so for the FC case, while less so for the RC case. A comparison of the two curves in Fig. 9 also shows that the results depend rather strongly on the intermolecular potential used. A less ambiguous quantity to calculate is the "hydration number", here defined as the number of water molecules within 5 Å from a given atom. Table VIII shows that the hydration number defined in this way is larger, as expected, in the FC case than in the RC case. Both potential models show the same trend, with a slightly larger hydration of the C1 and C2 atoms and a small but significant increase in the hydration number at the end of the carbon chain. That terminal carbon atoms are more hydrated than atoms in the middle of the chain is also supported by NMR relaxation data on fluorinated amphiphiles⁵³.

Another quantity of relevance for the interpretation of spectroscopic data is how the mobility of

water molecules is affected by the micellar surface. We have investigated the rotational motion of water molecules in successive shells surrounding the micelle. This has been done by calculating a time correlation function defined as

$$P_1(t) = \langle \underline{n}(t+\tau) \cdot \underline{n}(t) \rangle_t \quad (14)$$

where $\underline{n}$ is either the dipole vector or the H-H vector for a water molecule in one of the shells. In both cases three shells of a thickness of 2 Å and with inner radii of 10, 12 and 14 Å were chosen. The first shell between 10 to 12 Å is then expected to be most perturbed by the micelle, while the third shell between 14 to 16 Å should behave more or less as bulk water. What is found is that water molecules in the three shells behave very similarly, with approximately the same reorientational correlation times. The correlation times given in Table IX have been calculated from the logarithm of the corresponding tcf:s between 0.5 and 1.5 ps. The decay of the dipole vector tcf in the first shell (Fig. 10) is possibly slightly slower. The remaining differences are within the statistical uncertainties.

The translational diffusion of water molecules close to the micelle can be studied in terms of their propagators. Let us define a conditional probability, $p(k,t|k_0)$, to find a water molecule in shell k_0 at time t_0 and in shell k at time t_0+t . Such probabilities are quite easy to monitor and Fig. 11 shows three (self) propagators $P(k,t|k)$, $k = 1,2,3$, from the RC simulation. From these curves it seems that the translational motion in the first shell is as fast, or even faster, than in other shells. The above defined propagators can be obtained by integration from the corresponding macroscopic diffusion equation:

$$P(r,t|r_0) = \frac{2r}{(4\pi Dt)^{0.5}r_0} \sinh \left[\frac{rr_0}{2Dt} \right] \exp \left[-\frac{r^2+r_0^2}{4Dt} \right] \quad (15)$$

Fitting the solutions of Eq. 15 to the simulated curves in Fig. 12 by optimising the diffusion coefficient D gives 11, 8 and $7 \cdot 10^{-9} \text{ m}^2/\text{s}$ for the three shells, thus supporting the conclusion drawn directly from the curves in Fig. 11. Eq. 15 is not strictly appropriate since the micellar core should act as an excluded volume. However, Eq. 15 gives lower bounds for the optimised diffusion coefficients (the solution to the diffusion equation with an excluded volume can be found in ref. 50). The simulation diffusion coefficient of pure water⁴¹ is $6 \cdot 10^{-9} \text{ m}^2/\text{s}$, similar to the values stated above.

Consequently, we conclude that the rotational and translational motion of water molecules are not much affected by the presence of a micelle.

Figure 13 depicts the pair correlation function between the carboxyl carbon and the water oxygen. There is a strong first neighbour peak due to the electrostatic attraction between the carboxyl carbon and the water oxygen and between water hydrogens and carboxyl oxygens. In the FC case the peak is strong and even a second neighbour peak can be seen while in the RC case the peak is much weaker.

Counterion distribution

The ion distribution relative to the com of the micelle is shown in Fig. 14. In the RC case the concentration is almost uniform outside the micelle with only a small peak of ions that are electrostatically attracted to the charged head groups. In the FC case these ions form a much larger and broader peak. This may be further quantified by calculating the average number of ions inside a radius of 14 Å. In the FC case it is 10.7 while in the RC case it is 4.8. It is not straightforward to compare to a continuum model, since the choice of dielectric permittivity and micellar and ionic radii is ambiguous. However, with the micelle and ion radii equal to 10.5 and 1.5 Å respectively, and $\epsilon_r = 80$, the HNC approximation gives an average number of ions of 5.8 within 14 Å. However, this does not mean that the ions are bound to the carboxyl groups in the FC case. As seen from Fig. 15 there is a very strong first neighbour peak at about 2.25 Å in the carboxyl oxygen sodium ion pair correlation function. However, the area under the whole of this peak just corresponds to 0.25 ions. There is a minimum at 3.25 Å and a very broad maximum corresponding to a concentration that is less than twice that reached at large distances. In the RC case there is still less structure in this correlation function. Thus, there is no support for the concept of "bound ions" in the present simulations. Table VIII also contains the number of sodium ions within 5 Å from the carbon atoms. There is a clear difference between the FC and RC simulations, but, as for the hydration number, they show the same dependence on chain position. Contrary to the hydration number, the number of close ions decreases monotonically along the chain and is smallest for the methyl group. This agrees qualitatively with results from paramagnetic relaxation studies by Zemb and Chachaty⁵¹. The different behaviour of the hydration number and the number of close sodium ions can be understood from the fact that the alkyl chain has a lower dielectric permittivity than water. Thus, an alkyl chain is more likely to protrude into parts of the aqueous region that are devoid of ions.

The ion ion pair correlation function for the RC model, cf. Fig. 16, displays a first neighbour maximum at 3.2 Å and a second maximum at 10.7 Å. This unusual behaviour can be explained by the accumulation of counterions close to the micellar surface. Similar pair correlation functions have been obtained in continuum simulations of micellar solutions³⁰.

Summary

Results are sensitive to the choice of potential. The FC model gives a loosely organized micelle with considerable water penetration, while the RC model leads to a more compact micelle. Although the two micellar assemblies look quite different, they have several properties in common. The RC micelle is in better agreement with the common picture of a micelle. Also theoretical argument is in favour of this model, the main point being that the inaccuracies of the water model entails only inadequate screening of electrostatic interactions. The pseudoatom approximation and neglect of electronic polarisation is tantamount to a dielectric permittivity of 1 for the hydrocarbon phase. This will counteract the forming of micelles, as the dielectric screening reduces the repulsion between amphiphile head groups. As electrostatic interactions are overestimated, a model with reduced charges better corresponds to reality.

Both FC and RC models give broad radial distributions for the carboxyl carbons (although the FC one is twice as broad as the RC one). The standard deviations of the distributions do not vary much with chain position. This picture contradicts models where the head groups are fixed on a spherical shell or around a spherical shell with a certain spread.

The dynamics of the simulation is too fast, mainly due to deficiencies in the water model⁴¹, to be compatible with experiment. This includes rotational and translational diffusion of the micelle, reorientation rates of the hydrocarbon chains and rotational and translational diffusion of the water molecules. The water influences the motion of the amphiphiles, but the discrepancy between simulation and experiment is too large to be fully explained by this. For example, the decay of angular tcf:s has been observed to be too fast also for systems devoid of water⁵⁴. Since NMR measurements are used for comparison this may partly be due to inadequate treatment of the long time tails of the simulation tcf:s, but again this cannot account for the entire difference. A further contribution to the difference may arise from the dihedral potential. The relative energies of its minima are fairly well known, while those of its barriers are not. The error in barrier heights for butane may amount to 2.5 kJ/mol, or a factor of 2 in transition rate, and the errors in dihedral potential for larger hydrocarbons may well surpass that. Lastly, methyl and methylene groups were treated as pseudoatoms. While such an assumption may not infringe on the reproduction of equilibrium properties, there is no reason to expect it to reproduce dynamics correctly. In this context we advocate the use of dynamic data in the construction of molecular potentials.

The water close to the micellar surface behaves, dynamically, very similarly to bulk water, as represented by water at the edges of the simulation box and by simulations of pure water⁴¹. This contradicts the concept of "bound water", i. e., that water molecules outside large molecules move more slowly.

Despite uncertainties about potential and dynamics, the simulations reported here are based on a more accurate model than earlier statistical mechanical treatments, including other simulations. It

does not rely on arbitrary assumptions about radial distributions or bond directions. In consequence, the micelle is maintained by forces due to water-hydrocarbon interactions and not by *ad hoc* restraints. Thus, the structure of the micelle should be more realistic than that obtained from earlier models.

Acknowledgements. This work was supported by the Swedish Natural Science Research Council through generous allocation of computer time. The graphics program package SPACEFIL was kindly put at our disposal by R. Lindh. Interesting discussions with J. C. Eriksson, U. Henriksson, B. Lindman, E. Hansson and H. Wennerström are gratefully acknowledged.

References

1. J. W. McBain, *Trans. Far. Soc.*, **9**, 99 (1913).
2. F. M. Menger, *Acc. of Chem. Res.*, **12**, 111 (1979).
3. K. A. Dill & P. J. Flory, *Proc. Natl. Acad. Sci. USA* **78**, 676 (1981).
4. P. Fromherz, *Chem. Phys. Lett.*, **77**, 460 (1981).
5. K. A. Dill, *J. Phys. Chem.*, **86**, 1498 (1982).
6. F. M. Menger & D. W. Doll, *J. Am. Chem. Soc.*, **106**, 1109 (1984).
7. K. A. Dill, D. E. Koppel, R. S. Cantor, J. D. Dill, D. Bendedouch & S.-H. Chen, *Nature* **309**, 42 (1984).
8. F. M. Menger, *Nature*, **313**, 603 (1985).
9. K. A. Dill, *Nature*, **313**, 603 (1985).
10. J. A. Butcher Jr. & G. W. Lamb, *J. Amer. Chem. Soc.*, **106**, 1217 (1984).
11. B. Cabane, *J. Phys. (Orsay, Fr.)*, **42**, 847 (1981).
12. G. J. T. Tiddy, *Spec. Period. Rept. Nucl. Magn. Res.*, **6** (1976).
13. D. Bendedouch, S.-H. Chen & W. C. Koehler, *J. Phys. Chem.*, **87**, 153 (1983). D. Bendedouch, S.-H. Chen & W. C. Koehler, *J. Phys. Chem.*, **87**, 2621 (1983). J. B. Hayter and T. Zemb, *Chem. Phys. Letters*, **93**, 91 (1982).
14. B. Lindman, N. Kamenka, M.-C. Puyal, B. Brun & B. Jönsson, *J. Phys. Chem.*, **88**, 53 (1984).
15. See for example H. Wennerström & B. Lindman, *Phys. Rept.*, **52**, 1 (1979).
16. S. W. Haan & L. R. Pratt, *Chem. Phys. Lett.*, **79**, 436 (1981).
17. B. Owenson & L. R. Pratt, *J. Phys. Chem.*, **88**, 2905 (1984).
18. D. W. R. Gruen, *J. Phys. Chem.*, **89**, 146 (1985).
19. D. W. R. Gruen, *J. Phys. Chem.*, **89**, 153 (1985).
20. A. Ben-Shaul, I. Szleifzer & W. M. Gelbart, *Proc. Natl. Acad. Sci. USA*, **81**, 4601 (1984).
21. A. Ben-Shaul, I. Szleifzer & W. M. Gelbart, *J. Chem. Phys.*, in press. A. Ben-Shaul, I. Szleifzer & W. M. Gelbart, *J. Chem. Phys.*, in press.
22. A. Ben-Shaul & W. M. Gelbart, *Annual Rev. Phys. Chem.*, **36**,
23. J. M. Haile & J. P. McConnell, *J. Phys. Chem.* **88**, 6363 (1984).
24. M. C. Woods, J. M. Haile & J. P. McConnell, *J. Phys. Chem.*, in press.
25. J. Th. G. Overbeek & D. Stigter, *Rec. Trav. Chim.*, **75**, 1263 (1956)
26. C. Tanford, *Proc. Natl. Acad. Sci. USA*, **71**, 1811 (1974)
27. G. Gunnarsson, B. Jönsson & H. Wennerström, *J. Phys. Chem.*, **84**, 3114 (1980).
28. D. F. Evans, D. J. Mitchell & B. W. Ninham, *J. Phys. Chem.*, **88**, 6344 (1984).
29. P. Linse, G. Gunnarsson & B. Jönsson, *J. Phys. Chem.*, **86**, 413 (1982).
30. P. Linse & B. Jönsson, *J. Chem. Phys.*, **78**, 3167 (1983).
31. P. van der Ploeg & H. J. C. Berendsen, *J. Chem. Phys.*, **76**, 3271 (1982).
32. P. van der Ploeg & H. J. C. Berendsen, *Mol. Phys.*, **49**, 233 (1983).
33. W. F. van Gunsteren & M. Karplus, *Macromolecules*, **15**, 1528 (1982).
34. M. Margenau & N. R. Kestner, *Theory of Intermolecular Forces*, (Pergamon Press, New

York, 1969).

35. S. Engström, B. Jönsson & B. Jönsson, *J. Magn. Res.*, **50**, 1 (1982).
36. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren & J. Hermans, in *Intermolecular Forces*, B. Pullman(ed.) (D. Reidel, Dordrecht, 1981).
37. M. Neumann, *J. Chem. Phys.*, **82**, 5663 (1985).
38. O. Matsuoka, M. Yoshimine & E. Clementi, *J. Chem. Phys.*, **64**, 1351 (1971).
39. J. M. Caillol, D. Levesque, J. J. Weis, P. G. Kusalik & G. N. Patey, *Mol. Phys.*, **55**, 65 (1985).
40. O. Teleman & B. Jönsson, *J. Comp. Chem.*, **7**, in press (1986).
41. O. Teleman, S. Engström & B. Jönsson, manuscript in preparation.
42. B. Cabane, P. Duplessix & T. Zemb, personal communication.
43. *Handbook of Chemistry and Physics*, 50th ed., F-36, (The Chemical Rubber Co., Cleveland, 1968).
44. C. Hu & R. Zwanzig, *J. Chem. Phys.*, **60**, 4354 (1974).
45. P. J. Flory, *Statistical Mechanics of Chain Molecules* (Wiley, New York, 1969).
46. A. Abragam, *The Principles of Nuclear Magnetism* (Oxford University Press, Oxford, 1961).
47. H. Wennerström, B. Lindman, O. Söderman, T. Drakenberg & J. B. Rosenholm, *J. Am. Chem. Soc.*, **101**, 6860 (1979).
48. H. Walderhaug, O. Söderman & P. Stilbs, *J. Phys. Chem.*, **88**, 1655, (1984).
49. O. Söderman, H. Walderhaug, U. Henriksson & P. Stilbs, *J. Phys. Chem.*, in press.
50. H. S. Carslaw & J. C. Jaeger, *Conductions of Heat in Solids*, (Clarendon Press, Oxford, 1947).
51. T. Zemb & C. Chachaty, *Chem. Phys. Lett.*, **88**, 68 (1982). T. Zemb & C. Chachaty, *Chem. Phys. Lett.*, **91**, 240 (1982). See also Y. Chevalier & C. Chachaty, *J. Phys. Chem.*, **89**, 875 (1985).
52. T. Klason & U. Henriksson, in *Solution Behaviour of Surfactants*, vol. 1, K. L. Mittal & E. J. Fendler (eds.), 417-429, (Plenum Press, 1982).
53. J. Ulmius & B. Lindman, *J. Phys. Chem.*, **85**, 4131 (1981).
54. O. Edholm & J. Johansson, manuscript in preparation.

Tables

Table I. Lennard-Jones parameters for the interacting atoms and pseudoatoms. Net charges are given in units of the elementary charge $e = 1.602 \cdot 10^{-19}$ As, σ in Å and ϵ in kJ/mol. The numbering of atoms is explained in Fig. 1.

σ	1	2	3	4	5	6	7
1	3.207	3.296	3.341	3.029	2.667	3.189	0.000
2		3.920	3.920	3.118	2.667	3.279	0.000
3			3.920	3.163	2.667	3.323	0.000
4				2.851	2.667	3.011	0.000
5					2.667	2.667	0.000
6						3.166	0.000
7							0.000

ϵ	1	2	3	4	5	6	7
1	99.23	96.34	107.2	177.9	37.65	86.24	0.000
2		83.83	83.83	170.8	37.65	84.44	0.000
3			83.83	189.2	37.65	93.70	0.000
4				325.6	37.65	154.2	0.000
5					37.65	37.65	0.000
6						77.34	0.000
7							0.000

q	1	2	3	4	5	6	7
FC	0.270	0.000	0.000	-0.635	1.000	-0.820	0.410
RC	0.135	0.000	0.000	-0.3175	0.500	-0.820	0.410

Table II. Force constants and equilibrium values for bond lengths and bond angles. The numbering of atoms refers to Fig. 1.

Bond	Force constant	Equilibrium distance
1-2	3348 $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$	1.52 $\text{\AA}$
1-4	3766 "	1.22 "
2-2	3348 "	1.52 "
2-3	3348 "	1.54 "
6-7	4637 "	1.00 "
Angle	Force constant	Equilibrium angle
1-2-2	251.0 $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{rad}^{-2}$	1.927 rad
1-2-4	334.8 "	2.023 "
2-2-2	251.0 "	1.911 "
2-2-3	251.0 "	1.911 "
4-1-4	418.4 "	2.251 "
7-6-7	383.0 "	1.911 "

Table III. Thermodynamic and transport properties for the aggregate.

	FC	RC
Pressure (MPa)	-	152 ± 1
Mean radius of the micelle ($\text{\AA}$)	11.9 ± 4.0	9.5 ± 1.7
Translation diffusion coefficient (m^2/s)	$5.4 \cdot 10^{-10}$	$9.6 \cdot 10^{-10}$
Rotational diffusion coefficient of amphiphiles (s^{-1})	$2.0 \cdot 10^9$	$5.3 \cdot 10^9$
Rotational diffusion coefficient of the whole micelle (s^{-1})	$1.7 \cdot 10^8$	$4.7 \cdot 10^8$
Moment of inertia of the three ordered principal axes ($10^{-18} \text{ kgm}^2/\text{mol}$)	2.3 ± 0.2	1.0 ± 0.07
	2.0 ± 0.2	1.0 ± 0.07
	1.0 ± 0.04	0.67 ± 0.04

Table IV. Average distance from center of mass to different carbon atoms (C1-C8) and standard deviations in Å. Comparison is made with a "close packed" model, where all the C8 atoms are packed closest to the center followed by the C7 and so on, and to an "all trans" model.

Carbon atom	FC	RC	"close packed"	"all trans"
C1	11.9±4.0	9.5±2.3	9.1	13.6
C2	11.2±4.1	8.7±2.3	8.7	12.4
C3	10.8±4.3	8.0±2.3	8.2	11.1
C4	10.4±4.5	7.5±2.3	7.7	9.9
C5	10.1±4.8	7.2±2.3	7.1	8.7
C6	9.9±5.1	7.0±2.4	6.3	7.4
C7	9.9±5.3	7.0±2.4	5.3	6.2
C8	9.9±5.4	7.2±2.6	3.7	3.7

Table V. Fraction trans in the five dihedral angles.

Angle	1-4	2-5	3-6	4-7	5-8	Average
FC	0.50	0.51	0.65	0.54	0.47	0.53
RC (0-36 ps)	0.52	0.60	0.44	0.45	0.49	0.50
RC (36-72 ps)	0.55	0.50	0.51	0.52	0.40	0.50
Average	0.52	0.54	0.53	0.50	0.45	0.51

Table VI. Fraction chains with different number of gauche bonds (n_g).

n_g	0	1	2	3	4	5
FC model	0.05	0.13	0.43	0.24	0.16	0.00
RC model	0.02	0.15	0.30	0.41	0.11	0.02
Independent model with $E_g = 1.7$ kJ/mol	0.03	0.16	0.31	0.31	0.16	0.03

Table VII. Order parameters calculated from the orientation of the CH bonds of the different hydrocarbon groups calculated by three different methods, i) from the plateau of the tcf:s, given without sign as they are obtained as the square root of S^2 , ii) with respect to the local normal of the micelle and iii) with respect to the centre of mass to C1 vector.

FC model						
Chain position	2	3	4	5	6	7
i)	0.20	0.25	0.26	0.20	0.19	0.12
ii)	-0.13	-0.06	-0.01	+0.01	+0.03	+0.03
iii)	-0.18	-0.12	-0.09	-0.07	0.00	0.00

RC model						
Chain position	2	3	4	5	6	7
i)	0.17	0.20	0.19	0.15	0.19	0.11
ii)	-0.05	-0.01	-0.01	0.00	+0.03	0.00
iii)	-0.09	-0.07	-0.07	-0.07	-0.05	-0.03

Table VIII. Number of water molecules (n_w) and number of sodium ions (n_{Na}) within 5 Å from a given carbon atom (C1-C8) in the chain. The figures in parentheses in the third column have been obtained with a radius of 4 Å.

	FC-model		RC-model	
	n_w	n_{Na}	n_w	n_{Na}
C1	8.9	0.43	6.7(3.5)	0.17
C2	8.1	0.48	5.2(1.9)	0.13
C3	6.8	0.44	3.7(1.3)	0.08
C4	5.8	0.20	3.6(1.1)	0.06
C5	5.2	0.14	2.7(0.9)	0.05
C6	5.0	0.12	2.5(0.8)	0.05
C7	5.2	0.09	3.0(1.0)	0.05
C8	5.4	0.07	3.5(1.6)	0.04

Table IX. Rotational correlation times (ps) for the first Legendre polynomial of the dipole vector and H-H vector in three shells around the micelle. The shell radii are 10-12 , 12-14 and 14-16 Å. The statistical error is estimated to be ± 0.2 ps. The corresponding reorientational correlation times for bulk flexible SPC water are 2.0 ps for the dipole vector and 1.5 ps for the H-H vector⁴¹.

Shell	1	2	3
Dipole vector (FC)	2.3	2.0	1.8
H-H vector (FC)	1.3	1.2	1.3
Dipole vector (RC)	2.6	2.3	2.2
H-H vector (RC)	1.5	1.6	1.6

Figure captions

Fig. 1. Numbering of atoms in the system. Subscripts define the pseudoatom types of the interaction potential while superscripts give the chain position referred to in the analyses.

Fig. 2. Snap shot of micelle plus counterions a) after 65 ps of equilibration (1) and after 30 ps of production (2) using the FC model, b) after 95 ps of equilibration (1) and after 36 ps of production (2) using the RC model. Carbon atoms are hatched, oxygen atoms are blank and sodium ions are contoured.

Fig. 3. Total carbon density (mol/l) for the FC (dashed line) and RC (solid line) models as a function of distance from centre of mass of the micelle.

Fig. 4. Probability distribution $P_i = 4\pi r^2 \rho_i(r)$ in $\text{\AA}^{-1}$ for different carbon atoms as a function of distance from centre of mass of the micelle. Solid line refers to the RC model and dashed line to the FC model.

Fig. 5. $A(q)$ for the CH_3 density (normalized to 1 at $q=0$, cf. Eq. 7) for the FC and RC models compared to a gaussian fit to experimental results from small angle neutron scattering¹³. Also shown is the result for a uniform distribution. Rescaling of the RC distribution to a larger micelle (54 chains) gives perfect agreement with the experimental curve.

Fig. 6. Electron density (mol/l) as a function of distance from centre of mass of the micelle.

Fig. 7. Effective excess dihedral potential (kJ/mol) calculated from the average dihedral distribution.

Fig. 8. Angular time correlation function for the orientation of the C-H bonds with the RC model. The numbers shown are the chain positions of the CH_2 groups.

Fig. 9. The water concentrations for the FC and RC models in mol/l as a function of the distance from the center of the micelle.

Fig. 10. Time correlation functions for first order Legendre polynomials describing the orientation of the dipole and H-H vectors in water. Solid lines refer to the first shell and dashed to the third.

Fig. 11. Simulated water propagators $P(k,t|k)$ with k representing the three shells. Solid line refers to $k=1$ (innermost shell), dashed line to $k=2$ and dotted line to $k=3$.

Fig. 12. A comparison between the simulated propagator $P(1,t|1)$ and the corresponding quantity obtained from the diffusion equation (Eq. 14).

Fig. 13. The average water concentration in mol/l as a function of the distance from the carboxyl carbon for the FC (dashed line) and RC (solid line) models.

Fig. 14. Sodium ion concentration in mol/l as a function of the distance from center of mass of the micelle for the FC (dashed line) and RC (solid line) models.

Fig. 15. Sodium ion concentration in mol/l as a function of the distance from the oxygens of the carboxyl groups for the FC (dashed line) and RC (solid line) models.

Fig. 16. Sodium sodium pair correlation function (normalized in mol/l) in the micelle water system for the RC model.

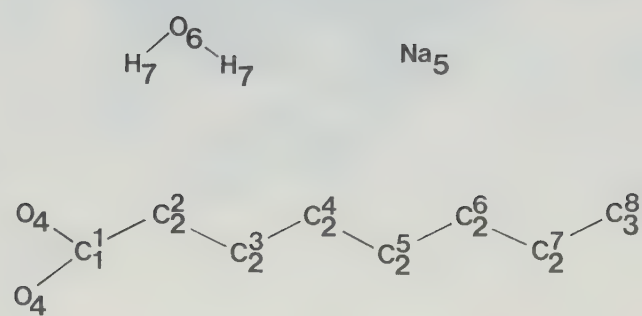


Fig. 1

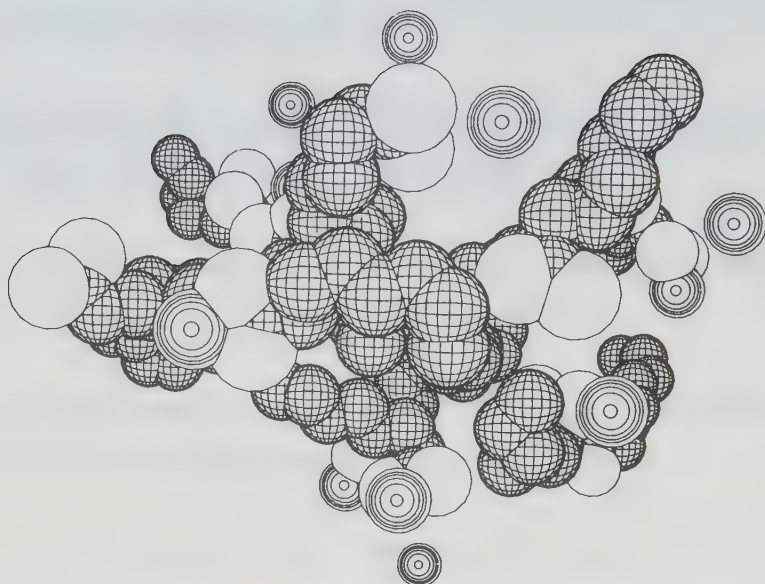


Fig. 2:a1

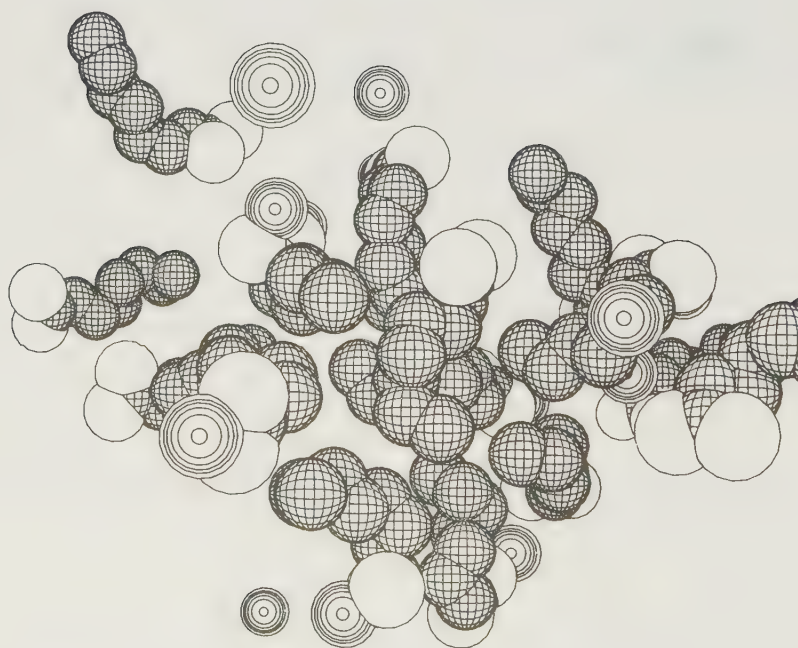


Fig. 2:a2

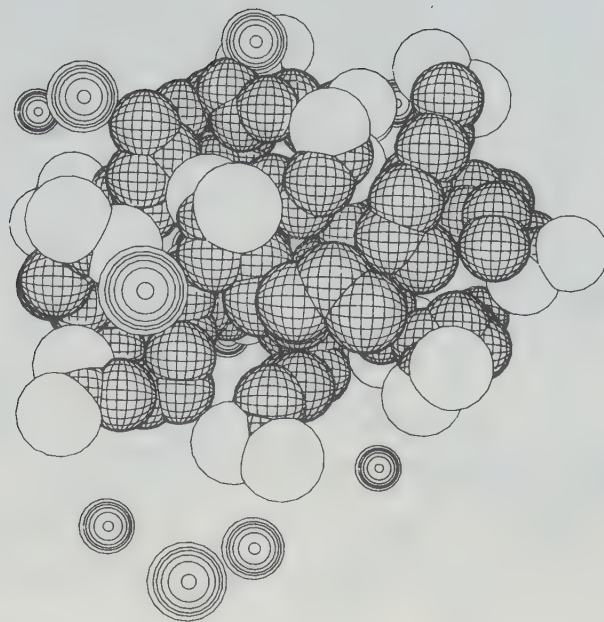


Fig. 2:b1

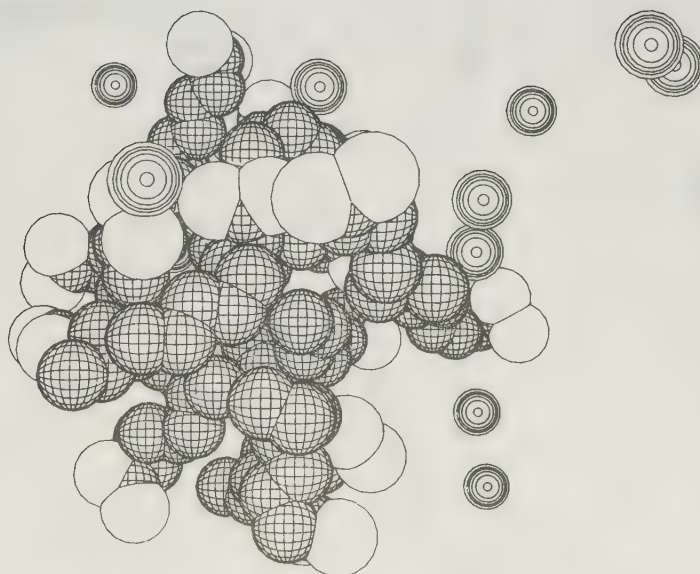


Fig. 2:b2

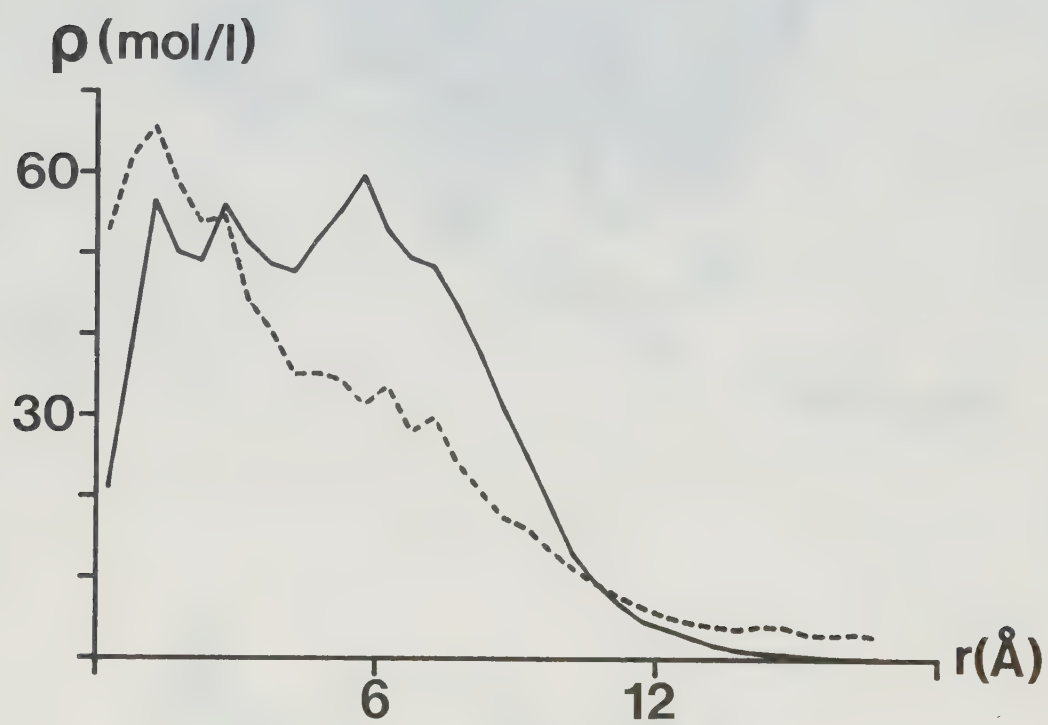


Fig. 3

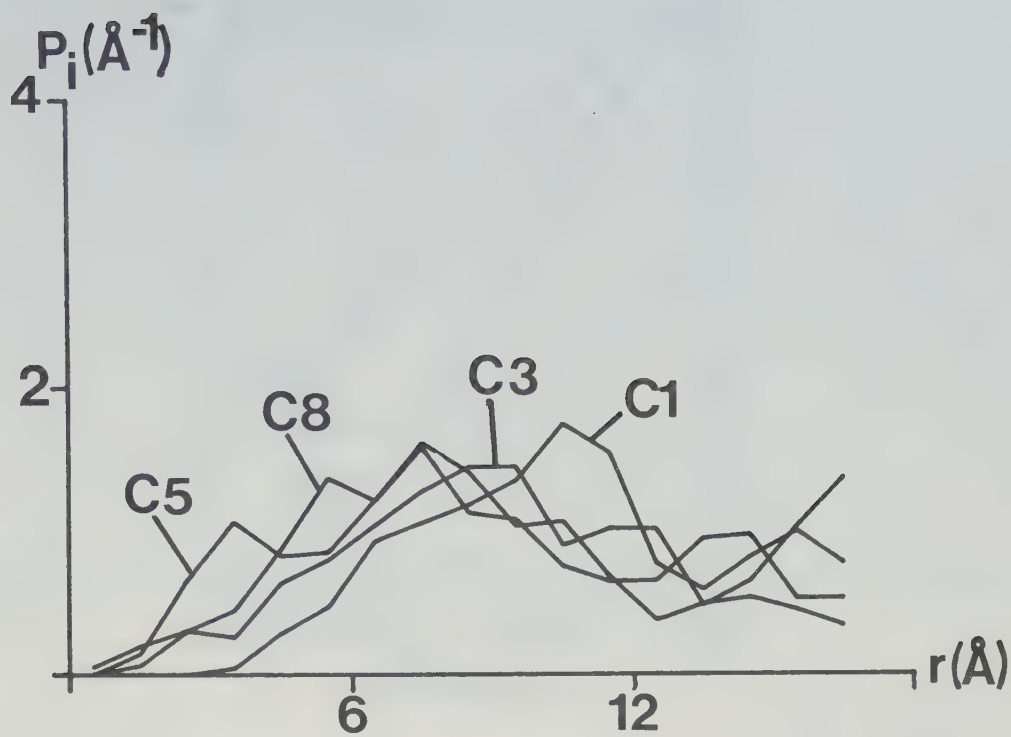


Fig. 4a

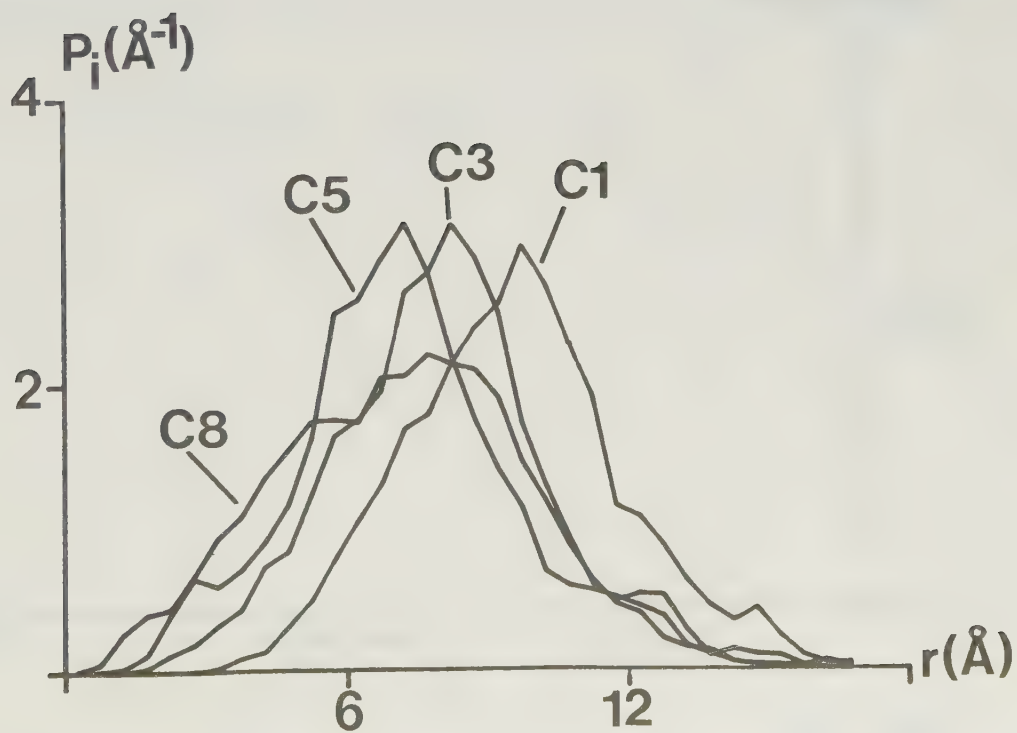


Fig. 4b

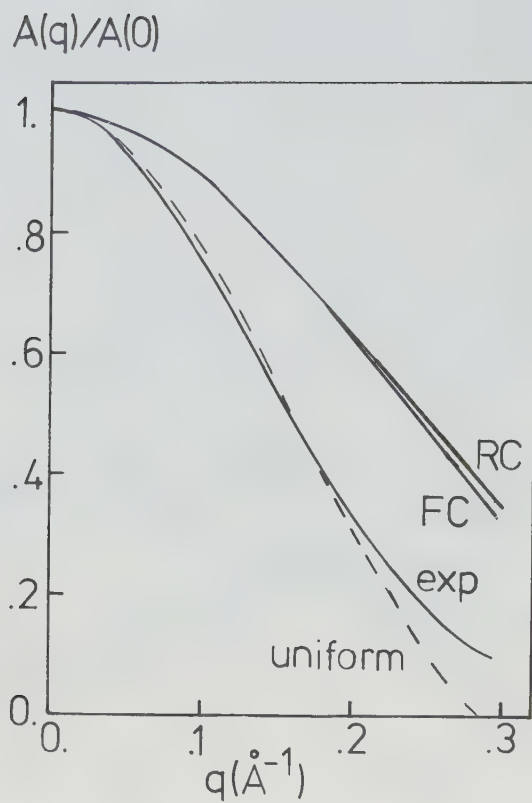


Fig. 5

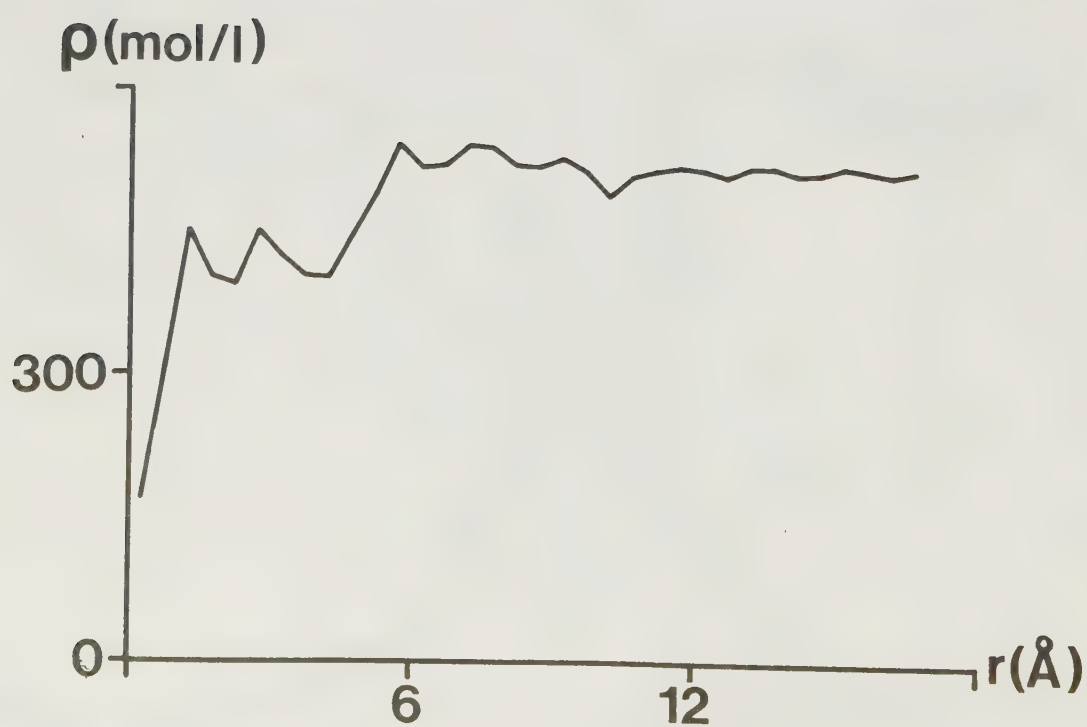


Fig. 6

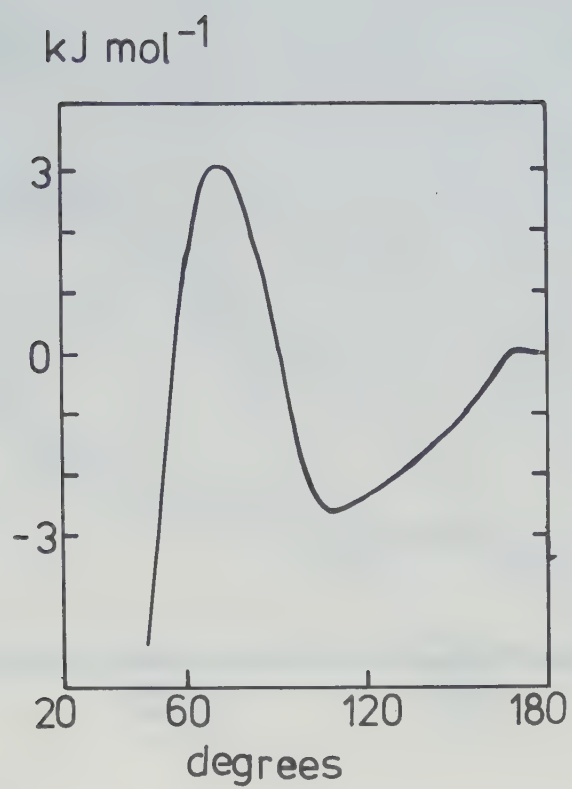


Fig. 7

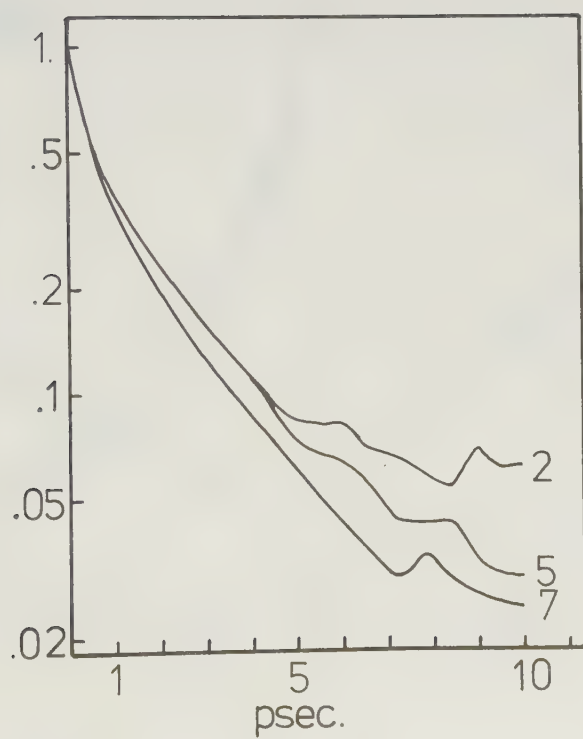


Fig. 8

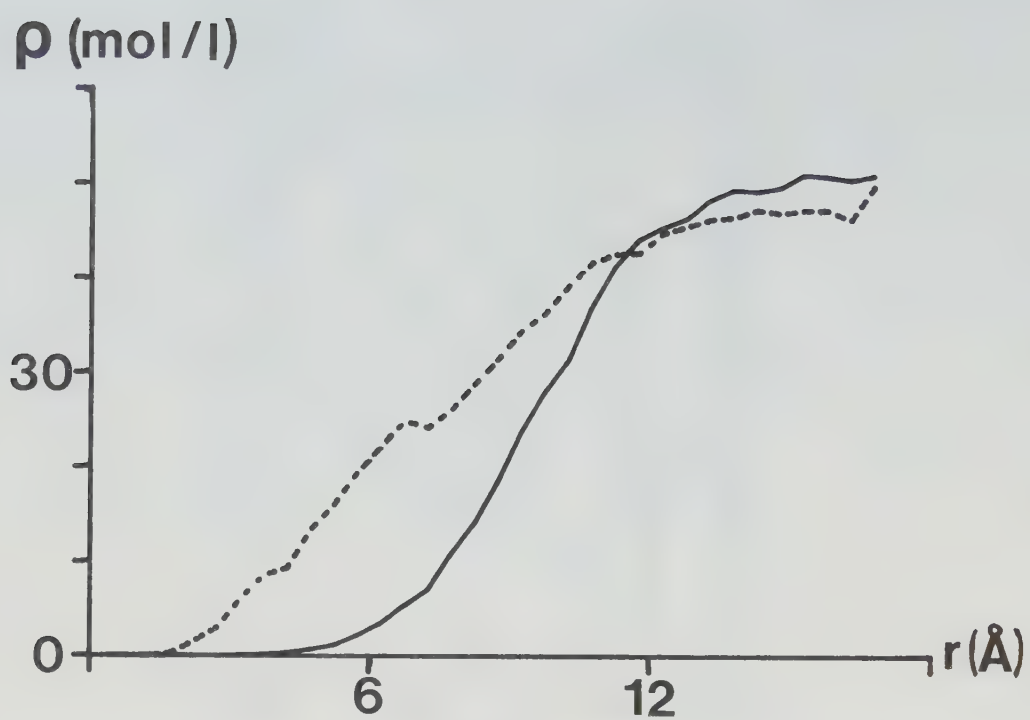


Fig. 9

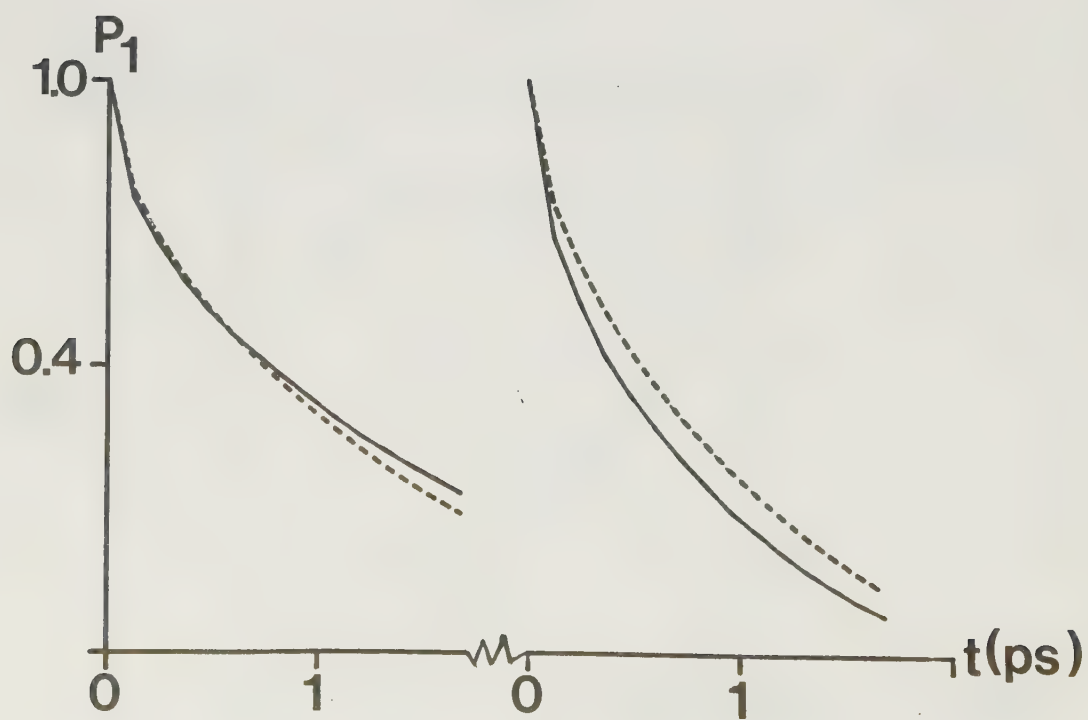


Fig. 10

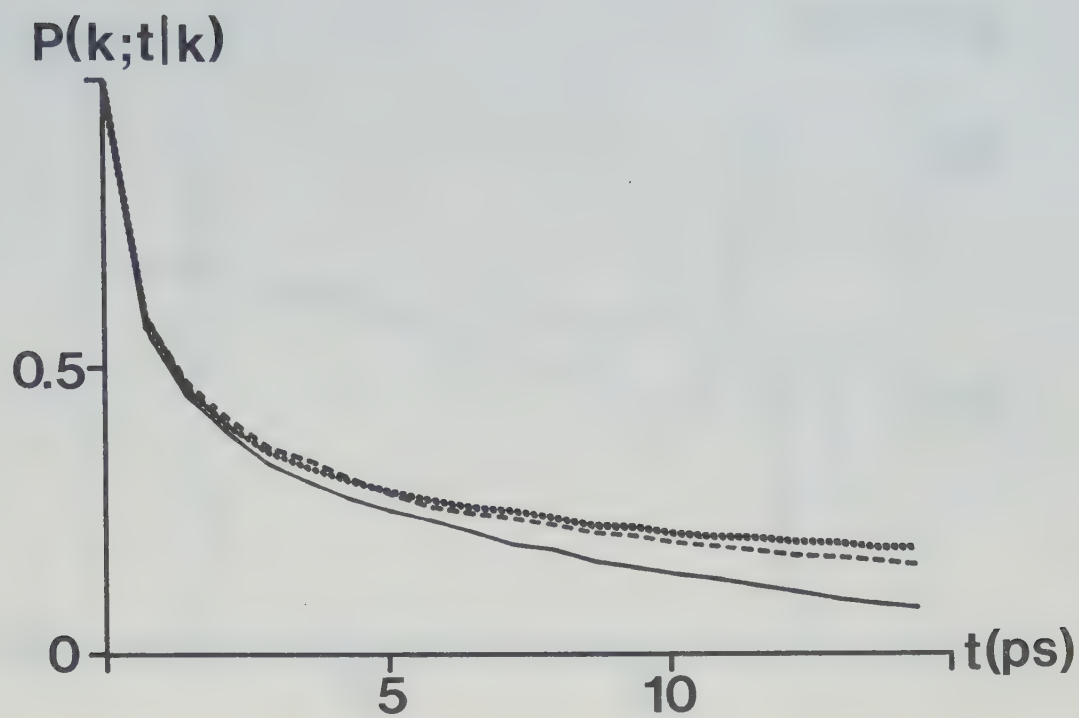


Fig. 11

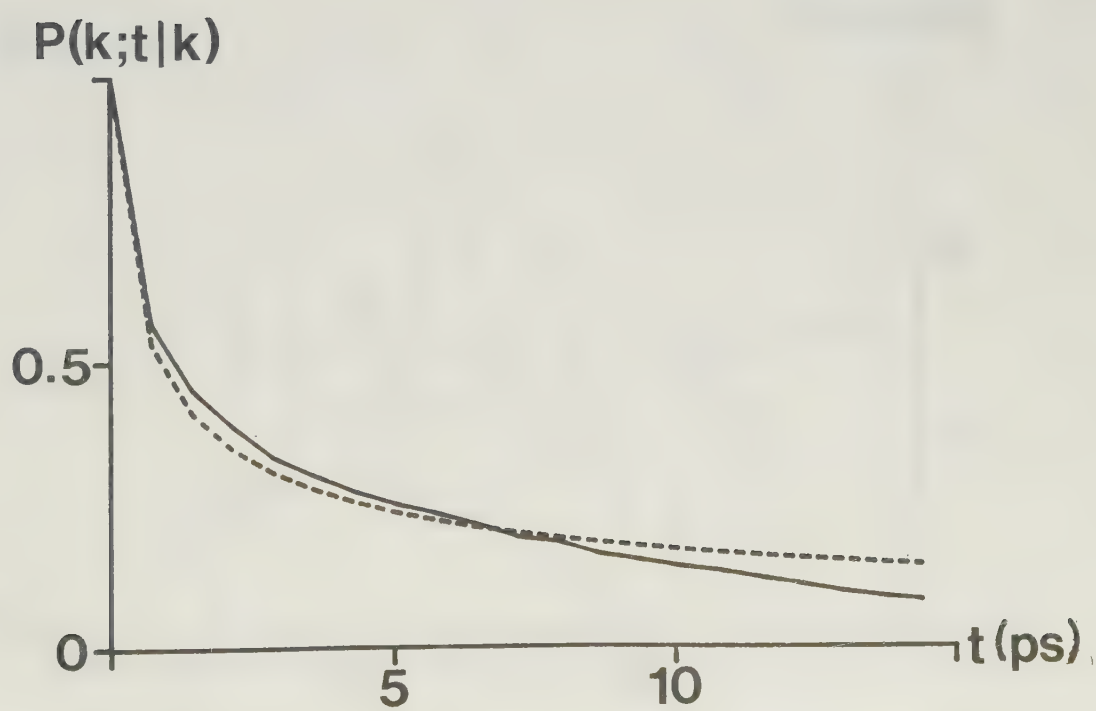


Fig. 12

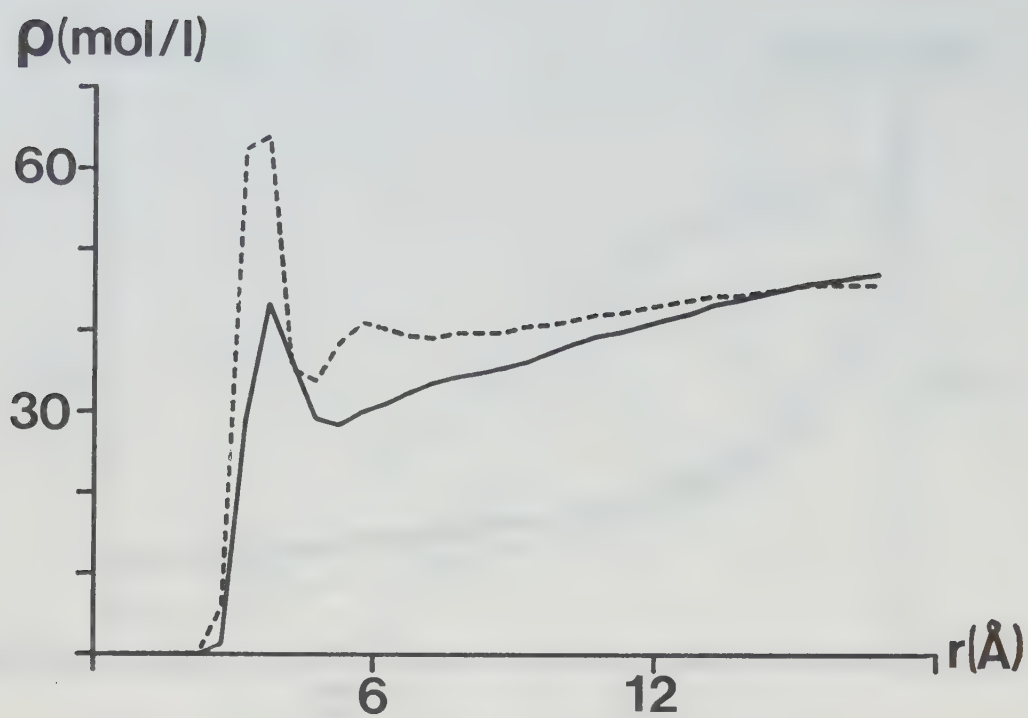


Fig. 13

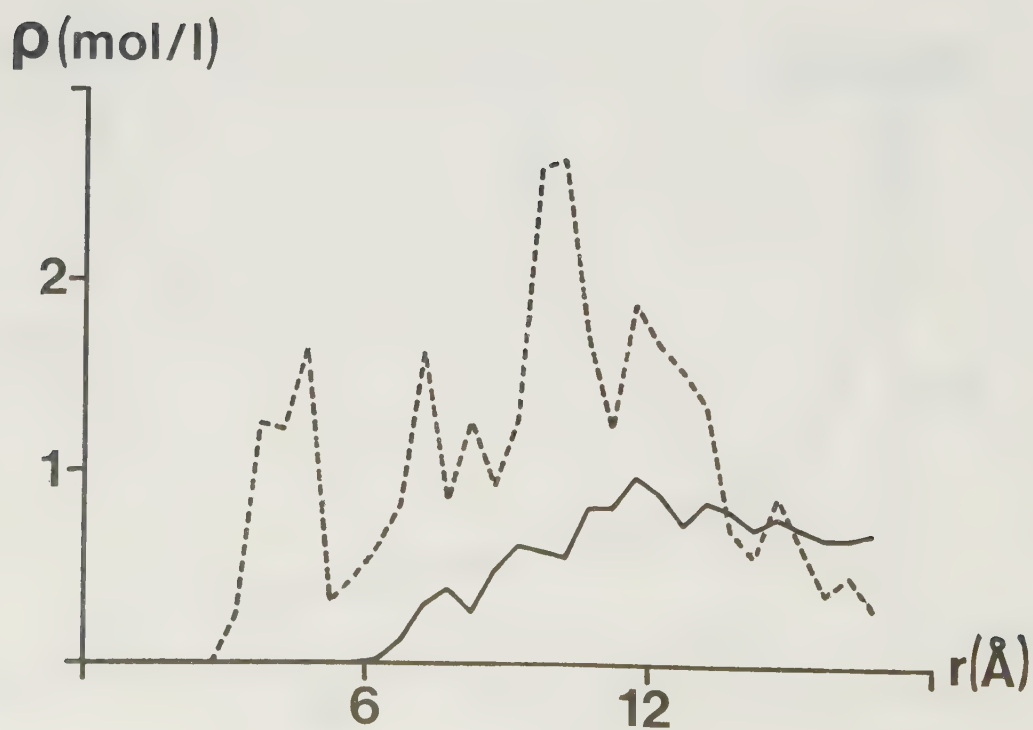


Fig. 14

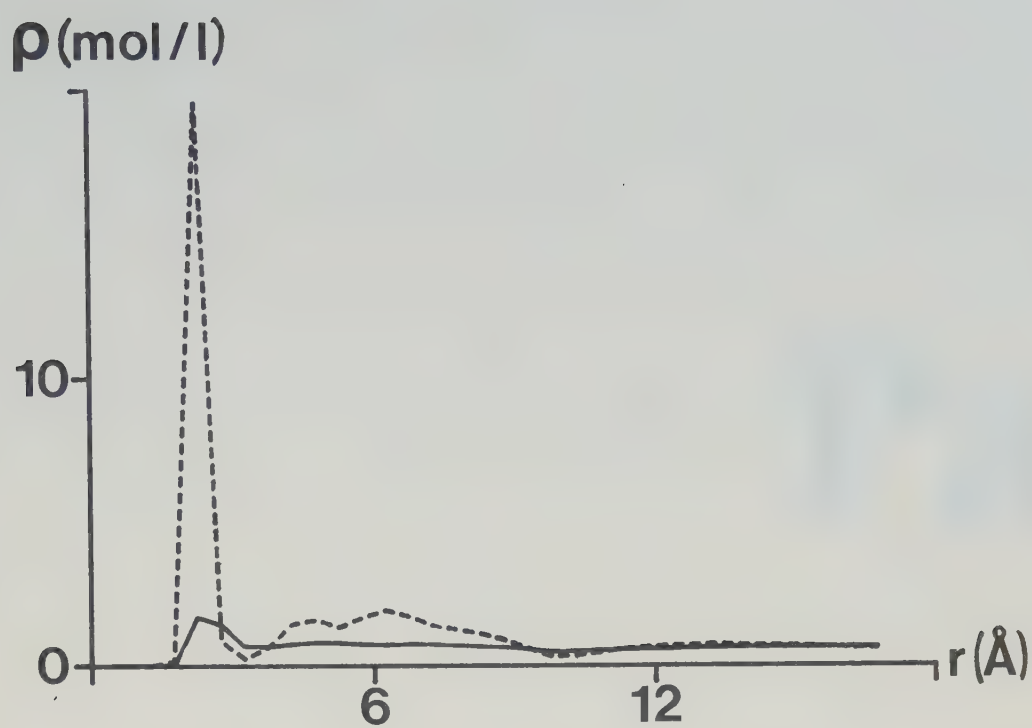


Fig. 15

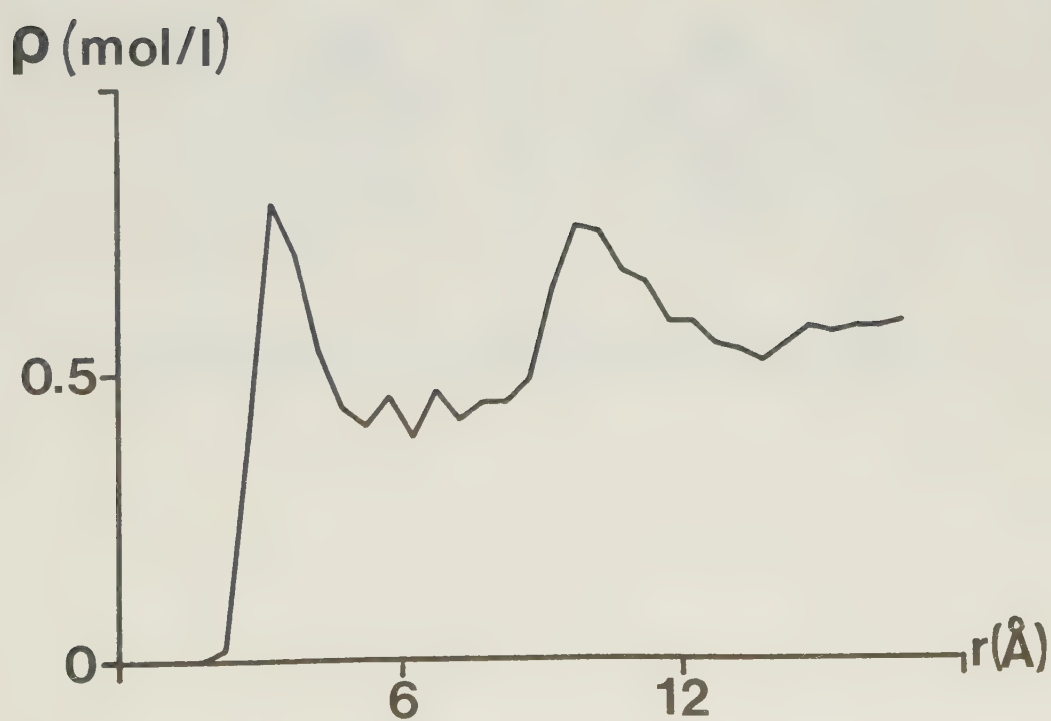
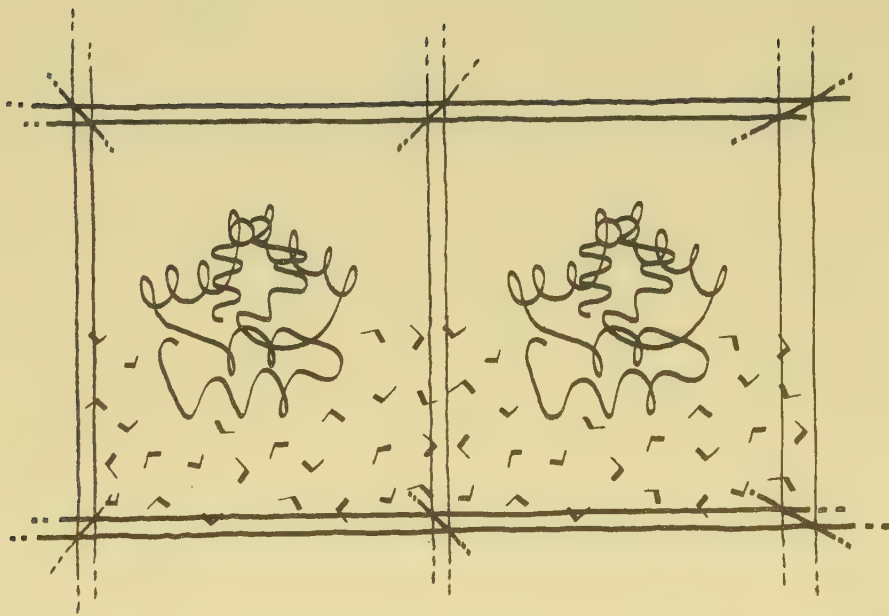


Fig. 16

Pa



Molecular Dynamics Simulation of Parvalbumin in Aqueous Solution

Peter Ahlström, Olle Teleman, Bo Jönsson and Sture Forsén

**Dept. of Physical Chemistry 2, Chemical Centre,
University of Lund, POB 124, S-221 00 LUND**

Abstract

Molecular Dynamics simulations of the Ca^{2+} binding protein parvalbumin are reported. The calcium saturated protein is simulated in an aqueous environment but also *in vacuo*. The Ca^{2+} free protein is studied as well, albeit only *in vacuo*. An X-ray diffraction structure is the starting point for all three simulations. Structural and dynamic properties are investigated and compared between the different simulations, each which lasted ≈ 100 picoseconds. The two *in vacuo* simulations show considerable structural changes relative to the initial coordinates. The entire protein contracts and its helix structures change. When simulated in aqueous solution, the protein still changes its structure relative to the crystal form, but less so than *in vacuo*, and helix structures largely remain intact. Dynamic properties of the protein are considerably altered on inclusion of water. Reorientational correlation times for vectors, both in the backbone and in side chains, change by orders of magnitude. This is the case for surface residues as well as internal. The calcium binding sites are accessible to the solvent and ligand exchanges are seen during the simulation. Removal of calcium ions causes global structural and dynamic changes. Internal protein hydrogen bonds are investigated. The number of hydrogen bonds is approximately the same in all three simulations, whereas their life time seems to be slightly longer in the *in vacuo* simulations. In general they are short-lived, with an average life time of a few picoseconds.

Introduction

During the last decade dynamic properties of biological macromolecules, in particular proteins and nucleic acids, have received considerable attention¹. This development has been stimulated by the recognition that the image of proteins and nucleic acids as rigid structures is highly incomplete and fails to capture the dynamic nature of biological processes. The binding of a substrate molecule to an enzyme, the chemical transformation of the substrate and its subsequent release must be associated with considerable conformational adjustments and changes in the mobility of many atoms. A much quoted example concerns the binding of an oxygen molecule to myoglobin or hemoglobin - the equilibrium structures of the deoxy form of these molecules leaves no room for the diatomic ligand to reach the buried heme binding sites²⁻⁴. The transfer of electrons in red-ox proteins will depend on vibronic couplings and fluctuations in the effective donor-acceptor distance and their relative orientation. The folding of a protein accompanying its emergence as a growing polypeptide chain from a ribosome is a relatively slow process involving a succession of transitory conformations while minimizing the free energy. Although our present understanding of molecular mechanisms of biological processes is far from complete, it appears that internal motions of proteins, relevant to their function, involve many degrees of freedom and cover a broad time scale ranging from picoseconds to seconds.

Experimental information regarding internal motions in proteins have been obtained through several methods. An early method, pioneered by Linderström-Lang⁵, is the exchange of hydrogen atoms between NH and OH groups of the proteins and solvent water. With the use of ¹H NMR methods, detailed information on the role of exchange of individual groups has been obtained in some cases, although the molecular details of the steps leading to exchange still remain obscure⁶⁻¹⁰. Particularly relevant are single crystal neutron diffraction studies, which show that most internal amide protons in a crystalline protein can be exchanged with overall rates comparable to those for the protein in solution¹¹⁻¹⁵. Studies of the dynamic quenching of fluorescent amino acids by oxygen and other molecules¹⁶⁻¹⁸ have indicated the interior of the proteins to be very accessible to small molecules.

Time resolved fluorescent anisotropy studies have been used to determine the motion of tyrosine and tryptophan side chains. Rotational correlation times in the order of a few hundred picoseconds or less have been measured^{19,20}. Perhaps the most detailed experimental information on internal motions in proteins has been provided by NMR spectroscopy. Recent developments in two-dimensional Fourier transform NMR have considerably simplified the assignment of observed resonances to atoms in individual amino acids in proteins. Measurements of relaxation rates (longitudinal and transverse) and nuclear Overhauser enhancements (NOE) probe internal motions in the time range from microseconds to fractions of nanoseconds. In addition, slower motions in the range 1 to 10⁻⁴ seconds - the flipping of some aromatic amino acids fall in this region - can be

followed from the line shape of NMR signals merging due to exchange of magnetic nuclei between different environments^{7,21-23}. In addition to the dynamic information obtainable from the techniques mentioned above, information on the mean square fluctuations of atoms from their equilibrium positions in crystalline proteins may be obtained from Debye-Waller factors²⁴. In the case of heme proteins, supplementary information on lattice disorder effects has been obtained from Mössbauer data^{4,24}.

A fundamental problem, common to all experimental techniques mentioned, concerns the detailed molecular nature of the dynamic processes observed. This particular problem is not inherent in some of the theoretical approaches to macromolecular internal mobility. The dynamics of a protein molecule can in principle be modelled from statistical mechanical simulation techniques. Provided that the potential energy of the macromolecule may be given as a function of the positions of all constituent atoms, the trajectories of all atoms is obtained by numerical integration of Newton's equations of motion. From the trajectories physical quantities of interest can be calculated. Starting with the pioneering molecular dynamics (MD) simulation by McCammon, Gelin and Karplus of the bovine pancreatic trypsin inhibitor (BPTI) with 58 amino acids²⁵⁻²⁷, this theoretical approach has attracted considerable attention. At the time of writing a considerable number of different macromolecular systems have been studied. For a review of the application of molecular dynamics simulations and related methods to proteins see for example ref. 28.

Most MD simulations of biological macromolecules have been done without explicit consideration of solvent molecules. One exception is the work by Swaminathan et al.²⁹, in which a comparison was made between a simulation of BPTI *in vacuo* and in a van der Waals solvent, the molecules of which having the density and molecular size corresponding to water molecules. van Gunsteren et al.³⁰ performed a short MD simulation of a unit cell of the BPTI crystal involving four protein molecules and 560 water molecules. The magnitude of the fluctuations calculated agree reasonably well with those earlier calculated for BPTI *in vacuo*, whereas the average structure was closer to the X-ray results. These studies have been taken to indicate that, while the motion of side chains at the protein surface will not be calculated correctly, solvent effects on fluctuations in the interior of proteins will generally be small. In particular, this has been assumed to be valid for the picosecond fluctuations²⁹⁻³².

In the present work we have attempted to investigate more closely the effects of solvent water on the dynamic properties of a Ca^{2+} binding protein. Parallel MD runs have been made *in vacuo* and in water. In the latter simulation a total of 2327 water molecules were included. Time correlation functions pertaining to different internal motions have been compared as have been radii of gyration, R factors, rotational diffusion and other properties.

The protein chosen for the present study, parvalbumin ($M_r = 11\,471$, see Fig. 1), belongs to a class of Ca^{2+} binding proteins that play an important role in living systems³³⁻³⁵. This class includes calmodulin, a ubiquitous protein vital to the Ca^{2+} dependent regulation of a wide variety

of cellular events, and troponin C, a protein present in all skeletal muscles and responsible for the Ca^{2+} dependent triggering of muscle contraction. These proteins are structurally homologous and the Ca^{2+} binding sites all feature the same helix-loop-helix structural arrangement - usually termed the "EF hand"^{34,36}. Parvalbumin has two EF hand type sites and is capable of binding two Ca^{2+} ions with high affinity ($K_A = 10^9 \text{ M}^{-1}$)³⁵. This protein therefore also offers an opportunity not earlier explored in MD simulations, namely to study the effect of removal of intrinsic metal ions on the dynamic properties of a protein. To investigate this a third MD simulation of parvalbumin with the two Ca^{2+} ions removed has been performed *in vacuo*.

Methods

Interaction potentials. The inter- and intramolecular interactions in a protein solution are immensely complicated and to a large extent inaccessible to quantum chemical calculations. *Ab initio* quantum chemical calculations yield intermolecular interactions between small molecules and ions as well as internal force field of small molecules. For example, the interaction between two water molecules has been calculated by Matsuoka et al.³⁷ and the resulting pair potential has been used in subsequent simulations of the liquid³⁸. Quantum chemical calculations of intermolecular potentials between water and amino acids have been reported by Clementi and co-workers³⁹⁻⁴¹. These potentials can after minor manipulations be combined so as to give the interaction between a water molecule and a peptide or a protein⁴². However, the accuracy of such a potential is difficult to assess. For example, the internal flexibility in the amino acid residues is not taken into account properly. This is the present limit of quantum chemical techniques, and the interaction within a protein or alternatively between different amino acids is still beyond reach.

In order to simulate protein solutions one has to resort to empirical or semi-empirical potentials. A number of different force fields have been proposed⁴³⁻⁴⁷ and used in protein simulations. The accuracy and validity of these empirical potentials are in general unknown. In most cases it is very difficult to find appropriate experimental data with which to test a potential. The confidence in the empirical potentials probably stems more from the analogy with simple liquids than from rigorous tests. An interesting question to ask, is how accurate potentials we do need. There is no straightforward answer, since it depends on, as in more simple cases, what properties we want to study. From our point of view the intermolecular potentials used so far seem to be rather crude and the protein simulations must be considered to be under development. This may sometimes be forgotten and results may not be interpreted with adequate caution.

So far, most protein simulations have been carried out *in vacuo* and the interest has been focused on small globular proteins with a minimum of charged residues. In reality, however, proteins do occur in solution and have a fair amount of charged residues, which may give rise to a

considerable net charge. We believe that electrostatic interactions play a vital role in biological processes and theoretical and technical problems associated therewith have to be dealt with. In many *in vacuo* simulations reported, the electrostatic interactions are attenuated by assuming a "distance dependent dielectric permittivity", supposed to mimic the solvent⁴⁷. The theoretical justification for such a procedure is not immediate and it has been subject to criticism⁴³. In a true molecular model, screening of electrostatic interactions should of course be due to the solvent. This places severe requirements on the dielectric behaviour of the water model used and it may preclude pair-wise additivity. One argument for the relevance of protein simulations *in vacuo* seem to be that, although the amino acid residues at the surface will behave differently with or without solvent, the interior should not be significantly affected. The validity of this assumption has to our knowledge not yet been verified.

In an attempt to approach these problems we have constructed an inter- and intramolecular potential based on site-site interactions, where a site can be either an atom or a group of atoms (pseudoatom). The only pseudoatoms are CH, CH₂ and CH₃ groups, and remaining hydrogen atoms have been treated explicitly. This leads to considerable savings in computing time for *in vacuo* simulations. However, when simulating a protein in aqueous solution, the aliphatic hydrogen atoms only represent a small fraction of the total number of atoms. Thus they can be included at only a marginal increase in computation. In fact, the aliphatic hydrogen atoms may play a significant role both for the protein structure and dynamics. For example, there is a non-negligible dipole moment, ≈ 0.4 D, associated with a C-H bond. In the present simulations we have refrained from using explicit hydrogen atoms in order to make more direct comparisons with previous simulation works.

A pair of sites i and j belonging to different molecules or separated by more than two bonds interact via a Lennard-Jones plus a Coulomb potential

$$U_{ij}(r) = 4 \epsilon_{ij} [(\sigma_{ij}/r)^{12} - (\sigma_{ij}/r)^6] + q_i q_j / 4\pi\epsilon_0 \epsilon_r r_{ij}. \quad (1)$$

The ϵ_{ij} and σ_{ij} are the usual Lennard-Jones parameters and q_i is the partial charge on site i . Equation (1) is the standard expression for the intermolecular potential in protein simulations. The only, but important difference is that ϵ_r is no longer distance dependent but set equal to one. The parameters have been taken from the literature^{46,48-50}, and the Lennard-Jones parameters were calculated from the Kirkwood-Slater formula⁴⁹ and from the assumption that the Lennard-Jones potential minimum corresponds to the sum of the van der Waals radii of the two atoms. The parameter set chosen may not be the most accurate one but should hopefully give a reasonable

description of short ranged interactions in the system. For example, the H₂O - amino acid potentials of Clementi and co-workers may well be a better choice than the now available empirical ones. For the present simulations, we have chosen not to use the quantum chemical potentials of Clementi et al., as their implementation is somewhat involved.

For the water-water interaction a number of different models exist, which all reproduce simple thermodynamic properties such as pair correlation functions, pressure and internal energy with reasonable accuracy. In the present work we have used a modification of the empirical simple point charge (SPC) model of Berendsen et al.⁵⁰. The main reason for this choice was simplicity, as the SPC model conforms to the interaction expression, eq. (1). Unfortunately, there has not, to our knowledge, been any determination of the dielectric permittivity for SPC water. One such simulation for the non-empirical potential of Matsuoka et al.³⁷ was reported recently, and its dielectric permittivity was found to be about 35 at room temperature⁵⁴, roughly half the experimental value for water. One likely reason for the too low value is the assumption of pair-wise additivity, neglecting nuclear and electronic polarisation. One may argue that SPC water should behave better, since it is an effective pair potential including many-body polarisation terms in an average way. However, this may not be the case, since the non-empirical potential of Matsuoka et al. has been fitted to a site-site model with a dipole moment of about 2.2 D, considerably larger than the experimental gas phase value, and close to the dipole moment of a SPC water molecule. Thus we expect SPC water to give a too low screening of ionic interactions.

The internal bond and bond angle vibrations were treated explicitly and assuming harmonicity the potential functions are given by

$$U_{\text{intra}} = \sum_{\text{bonds}} K_b (x_i - x_{i,\text{eq}})^2 / 2 + \sum_{\text{angles}} K_a (\theta_i - \theta_{i,\text{eq}})^2 / 2 \quad (2)$$

where x_i and $x_{i,\text{eq}}$ are actual bond length and equilibrium distance respectively and similarly for angles. The force constants K_a and K_b and the equilibrium values were taken from the literature^{48,50-53}. The interactions due to internal rotation were handled by means of periodic dihedral (torsional) potentials in the usual way⁴³.

$$U_{\text{int.rot}} = \sum_k \sum_{n=1}^3 C_{k,n} (1 - \cos[n\psi_k]) \quad (3)$$

where ψ_k is a dihedral angle and $C_{k,n}$ an interaction parameter. The parameters in eqs (1-3) are available as supplementary material from the publisher. Improper dihedrals (out-of-plane torsions) were not included as the bond angle potential takes care of these. No explicit hydrogen bond potential was used, instead it is assumed to be adequately described by Coulomb and Lennard-Jones terms.

Simulation technique. The Molecular Dynamics program used has been described in detail elsewhere⁵⁵ and here only a brief account of the technique will be given. The program is based on independent atoms or pseudoatoms and hence includes covalent interactions in terms of harmonic and periodic potentials. This usually requires a rather small time step when integrating Newton's equations of motion, due to rapid bond and bond angle vibrations. In order to circumvent this problem we have devised a double time step algorithm⁵⁵, where rapidly varying degrees of freedom are integrated with a small time step, while for slowly varying ones a larger time step can be used. In practice, bond and bond angle vibrations are integrated with $\Delta t = 0.2$ fs (fs = 10^{-15} s) while $\Delta T \approx 1-2$ fs is used for the remaining degrees of freedom. In the present simulations less than ten per cent of the total computation time was spent calculating the rapid motions, this is thus an efficient way to handle flexible molecules. It turned out that the Gear algorithm⁵⁶ is particularly suited for multiple time steps and we used a fourth order version of it.

Our algorithm differs from the usual approach, where bond lengths are kept fixed at their equilibrium value using constraints. There are several advantages with the double time step algorithm: i) The computer program is more easily "vectorized", ii) the algorithm is stable even with a very repulsive start configuration, and iii) the polarisation of bond lengths and angles is included, although only classically. The latter point may be important for the dielectric properties of the solution. A more detailed discussion of the consequences of including intramolecular vibrations in water is found in ref 57.

Two *in vacuo* simulations were performed, one with Ca^{2+} ions at the two binding sites and one devoid of calcium. Below, these simulations will be referred to as the VAC and APO simulations, respectively. A third simulation, AQ, was performed in aqueous solution with 2327 water molecules and three sodium ions so as to make the whole sample electroneutral. This corresponds to a protein concentration of 18 mM. Technical details for the three simulations are summarized in Table I. All three simulations were started from the X-ray coordinates (denoted F6-A) of Kretsinger and Nockolds⁵⁸. In the AQ simulation water molecules were placed on the bases of a primitive cubic lattice outside the protein. Three randomly chosen water molecules were exchanged for sodium.

Periodic boundary conditions were used in the AQ simulation, while the two others were performed in a cluster. All three simulations used a neighbour list technique and a spherical truncation of all interactions at $r_{\text{cut}} = 9$ Å, which leads to substantial computational savings in the

AQ simulation. The neighbour list does not render itself entirely to vectorization, but entails gather and scatter operations, which are quite time consuming. In consequence, a non covalent potential (eq. 1) of a more complex functional form only affects the total cpu time marginally.

Results and Discussion

Overall rotation and diffusion. Reorientational autocorrelation functions for vectors connecting two atoms have been calculated according to

$$C_j(t) = \langle P_j(\cos\theta(\vartheta, t)) \rangle_{\vartheta} \quad (4)$$

where $\theta(\vartheta, t)$ is the angle between the vector at a time ϑ and the same vector at a time t later, P_j is the Legendre polynomial of order j and the brackets indicate a time average. The j :th correlation time, τ_j , is defined through

$$\tau_j = \int_0^{\infty} C_j(t) dt \quad (5)$$

If $C_j(t)$ decays as a single exponential, i. e.

$$C_j(t) = A \cdot \exp(-t/T), \quad (6)$$

then the correlation time τ_j equals the time constant T . Sometimes it is more appropriate to approximate the correlation function with a sum of exponentials;

$$C_j(t) = A \cdot \exp(-t/\tau_{j1}) + B \cdot \exp(-t/\tau_{j2}) + C \cdot \exp(-t/\tau_{j3}) + \dots \quad (7)$$

where the different time constants may represent different processes, i. e. protein tumbling, side chain motions etc. There is often a rapid initial decay in a time correlation function, which is due to the atoms moving freely for a short while before colliding. In the idealized case, where $\tau_{j1} \ll \tau_{j2} \ll \tau_{j3} \ll \dots$, one can resolve the different exponents and in a given interval a single exponent governs the behaviour of the correlation function.

If the time correlation function obeys single exponential behaviour a correlation time τ_j can be estimated from

$$\tau_j = \frac{t_b - t_a}{\ln\{C_j(t_a)\} - \ln\{C_j(t_b)\}}, \quad (8)$$

where t_a and t_b are two different times.

Correlation times, τ_1 , for the first order Legendre polynomial have been estimated from the interval 1 to 8 ps of the correlation function from eq. (8). The time interval was chosen so as to avoid both statistical uncertainties at longer times and initial, rapidly decaying components.

In order to describe the overall rotation of parvalbumin in solution, $C_1(t)$ were calculated for several long vectors in the protein and their correlation times were estimated using eq. (8). The τ_1 values for vectors along the helices (cf below) range from 0.9 ns (helix A) to 2.5 ns (helix C) and for the vector between the calcium ions $\tau_1 \approx 1.7$ ns. Since the protein is flexible, it is difficult to define an unambiguous quantity describing the overall rotation. We will therefore use τ_1 for helix C as an estimate of the overall rotational correlation time.

The diffusion coefficient is formally defined as

$$D = \lim_{\Delta t \rightarrow \infty} \frac{\langle | \mathbf{r}(t+\Delta t) - \mathbf{r}(t) |^2 \rangle_t}{6\Delta t} \quad (9)$$

where $\mathbf{r}(t)$ is the position of centre of mass and D is the translational diffusion coefficient. The diffusion coefficient can then be calculated from the slope of the mean square displacement versus Δt in an interval where Δt is large compared to the correlation time (ζ^{-1}) for the velocity auto

correlation function. From the slope in the interval $\Delta t = 20\text{-}50$ ps $D = 2 \cdot 10^{-10} \text{ m}^2/\text{s}$ was obtained. The Einstein relation⁵⁹ $D = kT/m\zeta$ gives a correlation time for the velocity correlation function $\zeta^{-1} \approx 1.0$ ps. Thus the interval used is well inside the diffusion limit. The upper limit of the interval is set by the statistical uncertainties due to few points in the averaging.

From a hydrodynamic model, assuming the protein to be spherical, it is possible to obtain an effective radius as well a viscosity of water from the Stokes-Einstein formulae⁶⁰:

$$D = \frac{kT}{6\pi\eta r} \quad , \quad \tau_1 = \frac{4\pi\eta r^3}{kT} \quad (10)$$

Assuming that the Stokes radius is the same for translation and rotation we obtain

$$r_{\text{eff}}^2 = (1.5D\tau_1) \quad (11)$$

and

$$\eta^2 = \frac{(kT)^2}{54\pi^2\tau_1 D^3} \quad (12)$$

From the above values of τ_1 and D we find $r_{\text{eff}} = 9.1 \text{ \AA}$ and $\eta = 1.1 \cdot 10^{-3} \text{ Ns/m}^2$. The calculated radius, r_{eff} , is to be compared to the radius of $15.5 \text{ \AA}$ obtained from the radius of gyration. The viscosity is much larger than the approximate value of $0.24 \cdot 10^{-3} \text{ Ns/m}^2$ obtained for pure water using the same water model⁵⁷. We believe that the differences in radius and viscosity partly are due to our difficulties in estimating τ and D because of the shortness of the trajectory. The high viscosity may also be correct, since a real 0.02 M protein solution in general is highly viscous.

General structure of the protein. The deformation of the protein molecule was monitored by means of the radius of gyration (r_G), so-called R factors and changes in distances in the hydrophobic core. The radius of gyration is defined through

$$r_G^2 = \frac{\sum m_i r_i^2}{\sum m_i} \quad (13)$$

where m_i is the mass of the i :th atom, r_i is its distance to the centre of mass of the protein molecule, and the sums run over all atoms in the molecule. For a uniform sphere of radius r one can show that $r_G = 0.6^{0.5} \cdot r$. In the X-ray structure the radius of gyration of parvalbumin is 12.8 Å corresponding to a sphere radius of 16.5 Å. In all three simulations, r_G decreases, but more so in the two *in vacuo* simulations. At the end of the APO, VAC and AQ simulations $r_G = 11.5, 11.6$ and 12.0 Å respectively (corresponding to spheres with $r = 14.8, 15.0$ and 15.5 Å). The smaller radii of gyration in the *in vacuo* simulations in part reflect that the side chains tend to fold onto the surface to maximize favourable van der Waals interactions. Also the backbone contracts somewhat in all three simulations, as seen from the distances between all amide nitrogens of the backbone as a function of time.

It has been suggested⁶² that proteins should exhibit low frequency "breathing modes", that is, a periodic contraction and expansion of the whole protein with a frequency of $\approx 30 \text{ cm}^{-1}$. We have followed the time evolution of the radius of gyration, but have not been able to detect any significant periodic motions.

Comparison of the general structure of the protein with the X-ray structure was achieved by means of R factors as applied in the refinement of X-ray crystallography data⁶³. The R factor for a configuration was taken to be the minimum of

$$R = \left[\frac{\sum_{i=1}^N m_i |r_{i,a} - r_{i,X}|^2}{\sum_{i=1}^N m_i} \right]^{0.5} \quad (14)$$

where m_i is the mass of atom i and $r_{i,a}$ and $r_{i,X}$ its coordinate in the actual configuration and the X-ray structure, respectively. Minimization was performed with respect to interconfigurational position and orientation. The R factors for the backbone atom positions compared to their positions in the X-ray structure have been calculated every 10 ps for the trajectories (see Table II). The mean R factors from the two *in vacuo* simulations are about 3.2 Å and from the AQ simulation 2.5 Å. However, in all three cases and especially in the VAC simulation they tend to increase slightly, possibly indicating a continuous change in conformation. Many MD simulations of proteins report root mean square fluctuations around atomic average positions, that is $r_{i,X}$ in eq. (14) is replaced

by the average coordinate obtained during the simulation. These values, proportional to X-ray temperature factors, are generally in the order of 1 Å, depending on the protein under study as well as the length of simulation. If the protein undergoes large structural changes, then both R factors and "temperature factors" will increase with time. In most simulations this seems to be the case. For example, from a simulation of a BPTI crystal, van Gunsteren et al.³⁰ report R factors to increase from 1.44 Å after 8 ps to 2.11 Å 11 ps later. Similarly, Levitt⁶⁴ report R factors of 1.6 and 2.5 Å for C_α atoms and all atoms, respectively, from a 132 ps vacuum simulation of BPTI.

Another method to estimate the differences between the general structure in the three simulations is the distances between different residues. We have chosen to look at the mean distances between the centroids of the phenylalanines situated in the core in the X-ray structure, i. e. all phenylalanines except for Phe 2 and Phe 57, see Table III. It shows that the structure differs significantly between the different simulations and between each of the simulations and the X-ray structure. The results clearly show that the protein undergoes large changes relative to the initial X-ray coordinates. These results and also the R factors reported above may be due to artefacts in the potential function or due to different environments felt by the protein in a crystal, *in vacuo* (with or without Ca²⁺) and in solution. This is a question of paramount interest, which unfortunately we cannot answer at present. It is interesting to note the large differences between the *in vacuo* and the AQ simulations, clearly demonstrating the important role played by the solvent in determining the protein structure. The phenylalanines investigated in Table III are all situated in the core of parvalbumin, hence the structural influence of solvent is not only limited to the exterior part of the protein.

Backbone. It has been suggested that α-helices and β-sheets of a protein act as "springs" and show low frequency (5-10cm⁻¹) collective stretching vibrational modes⁶⁵. To investigate these hypotheses vectors parallel to the helices were defined between atoms at helix endpoints. The lengths of these vectors were calculated as functions of time and Fourier transformed. From the spectrum it was not possible to deduce the presence of slow stretching modes in any of the six helices, i. e. stretching modes seemed to be stochastic in all three simulations. In other *in vacuo* simulations⁶⁶ large scale vibrational modes have been detected. The protein investigated by Åqvist et al.⁶⁶ has a domain structure, whereas parvalbumin is a globular protein with no separable domains. However, the solvent considerably affects dynamics all over the protein and may have a damping effect on low frequency modes. Long range electrostatic interactions, cooperative in character, may also strongly affect the motional behaviour of large fragments in a protein.

The structure of a polypeptide chain can be specified by dihedral angles Φ and Ψ . (Φ_i is the angle between the planes C_{i-1}-N_i-C_{αi} and N_i-C_{αi}-C_i; Ψ_i the angle between the planes N_i-C_{αi}-C_i

and $C_{\alpha i}-C_i-N_{i+1}$.) We calculated the mean of these angles for all residues and plotted Ψ_i against Φ_i for each simulation in Ramachandran maps (Fig. 2). In the AQ simulation the points are more clustered than in the VAC and the APO simulations. The density is especially large in an area around $\Phi \approx -70^\circ$ and $\psi \approx -40^\circ$, close to the α -helix angles ($\Phi = -57^\circ$ and $\psi = -47^\circ$). In the VAC and APO simulations the spread of points may indicate a beginning denaturation of the α -helices. In all three simulations the area with higher density is moved towards lower Φ and higher ψ values compared to X-ray data.

The standard deviations were calculated for the dihedral angle distributions (Fig. 3) in order to investigate the torsional flexibility in different parts of the protein. In most cases the standard deviations due to fluctuations are 10-15 degrees, whereas structural changes are accompanied by larger deviations. In the AQ simulation this occurs near the N-terminal end of the chain and between residue 33 and 34, i. e. at the end of α -helix B, whereas the fluctuations in the α -helices are small. In the VAC simulation large deviations occur at the A helix, and between the A and the B helix and at residue 92 and 93, which are situated near the calcium loaded EF site. The largest deviations, nearly 90° , occur between the D and E helices. These fluctuations should not be interpreted as a high structural mobility, but are due to structural changes taking place during the simulation. In the APO simulation the backbone is very much affected by the absence of calcium, especially near the EF site where several residues show considerable deviations. The small changes at the CD site are discussed below.

The dynamics of the backbone was investigated by means of time correlation functions and correlation times (τ_1). These were calculated for vectors between α -carbons in neighbouring amino acid residues. Typical correlation times in the VAC simulation are between two and ten times (in some cases even more) longer than those of the AQ simulation. However, there are also some examples of correlation times being longer in the AQ simulation than in the VAC simulation. The results from the APO simulation give τ_1 values that in general are closer to the values from the VAC simulation.

The longest correlation times (in the order of nanoseconds) in the VAC simulation were obtained around the N-terminal end of the C-helix. Also in the APO simulation the longest correlation times were obtained in the C helix. In the AQ simulation the longest correlation times (about 1 ns, corresponding to half the overall rotational correlation time of the protein) were obtained in the region between the D and E helices. We note that parts with short correlation times normally show large standard deviations in the dihedral angles. This means that the correlation

times quoted in part may reflect structural changes rather than the dynamical behaviour. However, we do not think this obscures the interpretation significantly, since we find the most rapid motion in the AQ simulation where the structural changes are least.

Hydrogen bonds. Since no explicit hydrogen bond potential was used, it is of interest to investigate the behaviour of hydrogen bonds in the protein. The definition of a hydrogen bond is not trivial and has to be operational. To investigate possible geometric criteria, Reimers and Watts⁶⁷ used a local mode technique to obtain the vibrational spectrum of liquid water. The vibration frequencies were then correlated to a number of geometric properties, such as distances and angles, for intermolecular hydrogen-oxygen pairs constituting potential hydrogen bonds. From the knowledge that hydrogen bonds are more likely to be formed where vibrational frequencies are shifted to lower values, they found no simple and stringent geometric condition for the formation or existence of hydrogen bonds. Below, we will interpret simulation geometries in terms of hydrogen bonds, but we are aware that this designation is to be regarded with some caution.

A list of all polar hydrogens (i. e. those hydrogens that are taken into account explicitly in the simulation) and a list of all possible acceptors of hydrogen bonds (i. e. polar oxygens and nitrogens) were constructed. Water hydrogen bonds were not considered in this analysis. The distances between all possible combinations were calculated throughout the period of analysis. We chose to use an asymmetric criterion for hydrogen bonds such that a bond is considered to be formed when the distance between hydrogen and acceptor becomes less than some distance d_s and to cease when the distance becomes larger than d_e , where $d_s < d_e$. The reason for this is that encounters at long distance may not form any hydrogen bond. On the other hand a hydrogen bond already formed may vibrate and thus occasionally be very long without actually breaking. With $d_s = 2.1 \text{ \AA}$ and $d_e = 2.3 \text{ \AA}$ ($d_e = 3.5 \text{ \AA}$ in both cases) we obtained the results shown in Table IV. We note that hydrogen bonds in the protein are rather short-lived - especially in aqueous solution - and most hydrogen bonds lasted less than ten picoseconds. This is consistent with the properties of pure water. At room temperature it takes a water molecule $\leq 10 \text{ ps}$ to diffuse over a distance comparable to a molecular diameter. This can be taken as an upper limit for the hydrogen bond life time. It is assumed here that the diffusing species in water is a single water molecule.

Calcium binding sites. In the X-ray structure, the calcium ion in the CD site is liganded to six protein ligands and that in the EF site to seven protein ligands and one crystal water. In order to characterize calcium ligandation we calculated the distances between the calcium ions and all possible ligands. These were regarded to be ligands if closer to the calcium ion than $3.0 \text{ \AA}$. The results from this analysis are shown in Table V.

The crystal water from the X-ray structure was not included in the *in vacuo* simulations. No protein ligand was released during the VAC simulation, but further carboxyl oxygens became

liganded. In the CD site the calcium acquired two new ligands during the equilibration, O₈₂ from Asp 53 and O_{ε1} from Glu 62. At the EF site O₈₁ from Asp 90 arrives after 7 ps of analysis. These three new ligands all belong to carboxyl groups where the other oxygen atom was already a ligand, thus there are no drastic changes in the ligandation. We believe that this double ligandation of -COO⁻ groups is due to artefacts in the potential function. It would probably disappear if Ca²⁺ was allowed to polarise the diffuse electron cloud surrounding the -COO⁻ group. Similar effects have been seen in simulations of CaEDTA complexes⁶¹. The radial distribution functions for both sites show sharp peaks at 2.3 Å (Fig. 5). The ligands are not very mobile, the correlation times τ_1 for vectors in the liganding residues are with a few exceptions in the order of nanoseconds (see Table VI).

In the APO simulation, of course, the picture is quite another. In the absence of calcium ions mobility is enhanced and structure is lost. The only exceptions are a few ligands in the CD site which are about as immobile as in the VAC simulation (see Table VI). This is probably due to a neighbouring positively charged lysine (Lys 54), which stabilizes the negatively charged ligands.

The results from the AQ simulation differ in many respects from those of the VAC simulation. Crystal waters are present from the beginning and ligands exchange with water molecules from the solution. A water molecule enters the EF site after a few picoseconds of analysis and remains. The crystal water leaves the EF site after 10 ps. In the CD site a water molecule enters during the equilibration and remains. The protein ligands are less tightly bound to the calcium in water than *in vacuo*. Several protein ligands leave the calcium ion for a short while but return again. So is O_{ε1} in Glu 101 released after 24 ps of the analysed trajectory but returns 20 ps later. The calcium ions loose one protein ligand each compared to the X-ray structure. In the CD site, O_γ of Ser 55 departs after less than ten picoseconds of equilibration and in the EF site the peptide oxygen of Lys 96 leaves after 59 picoseconds of analysis. Further changes of ligands are frequent and are specified in Table V.

An interesting result is that coordination numbers are equal (≈ 8) at both calcium binding sites during most of the AQ and VAC simulations, whereas they are 6 at the CD site and 8 at the EF site in the X-ray structure. Other MD simulations give coordination numbers of ≈ 9 ⁶⁸ or ≈ 7 ⁶¹ for a calcium ion in water and ≈ 6.7 for a calcium ion in a CaEDTA²⁻ complex⁶¹.

The dynamics of protein ligands differ between the three simulations. This can be seen from the estimated τ_1 values for vectors between C_β and the carboxyl carbon in the non-exchanging ligands. These values are shown in Table VI (see also Fig. 7). From the table it is clear that

mobility in general is enhanced in the AQ simulation compared to the VAC simulation.

Side chains. Time correlation functions, $C_1(t)$, and an estimate of corresponding correlation times from the interval 1-8 ps were computed for the vector between the two outermost non-hydrogen atoms in all side chains except for alanines and glycines (Fig. 6). From a comparison of results from the two *in vacuo* simulations, one notices changes in correlation times all over the protein upon removal of the calcium ions. Mostly, the mobility is enhanced but in some cases there are changes going in the opposite direction.

A tendency valid for all types of residues is that correlation times are shorter in aqueous solution than *in vacuo*. This is very pronounced for charged and polar side chains at the surface, such as glutamic and aspartic acids, lysines and serines (see also Fig. 7), but nearly as pronounced for the hydrophobic phenylalanines in the core of the protein. Therefore we have analysed the behaviour of phenylalanines more carefully.

Time correlation functions have been calculated for the vector between $C_{\delta 1}$ and $C_{\delta 2}$ (transverse to the ring bond axis), denoted DD, and the vector between C_{γ} and C_{ζ} (approximately parallel to the ring bond axis), denoted GZ. The sums of these correlation functions are shown in Figure 8. It is clear that the effect of water is large - about as large as the major changes induced by the removal of calcium ions. We also estimated correlation times as described above and obtained for the DD vector $\tau_1 \approx 190$ ps in the VAC and 28 ps in the AQ simulation and for the GZ vector 250 ps and 109 ps respectively. One could argue that the difference between the simulations is due merely to those phenylalanines situated at the surface of the protein. To clarify this, we also plotted the individual time correlation functions. One example, Phe 30, situated in the core of the protein, is given in Fig. 9. In the AQ simulation there is a rapid decay of the correlation function in the beginning. This is probably due to the protein atoms being less densely packed than *in vacuo*, and hence having a longer mean free path. Another feature is the large difference between DD and GZ correlation times. This implies a considerable rotational motion of the ring around the C_{γ} - C_{ζ} axis.

This can also be seen from the angle between the plane defined by C_{α} , C_{β} and C_{γ} , and that by C_{γ} , $C_{\delta 1}$ and $C_{\delta 2}$. It shows fast rotational motions spanning about 30° and some occasional larger turns. Phe 30 has an increased mobility in water compared to *in vacuo*, DD correlation times are 374 ps in the VAC and 20 ps in the AQ simulation. Similar results are obtained for several other internal phenylalanine residues as well, which shows that the presence of water does indeed influence the dynamics not only of external parts of the protein but also of the interior.

Conclusion

The protein structure deviates from its crystal form in all three simulations. This is, of course, a consequence of the potential, but whether it is merely an artefact or a representation of the real conformation remains to be shown. It is certain, though, that the parvalbumin molecule has to be more or less flexible in order to function. Simulation confirms this picture of a fairly fluid molecule.

Water is paramount, and its presence affects dynamics and structure of the entire protein. Large effects on dynamics are found not only at the surface but also in the interior. The protein structure in aqueous solution more resembles the crystal forms than do the *in vacuo* ones, as manifested by surface side chains and backbone dihedral angles.

Ligandation of the calcium ions is slightly different for the two binding sites in the crystal form. This difference is less pronounced in simulation. Exchanges of water ligands do occur, which implies that binding sites are solvent accessible as, indeed, they have to be. Removal of the calcium ions affects local dynamics in all parts of the molecule. The protein is thus capable of fast, global information transfer, a prerequisite of the strong calcium binding cooperativity found from experiment⁶⁹.

The overall translational and rotational diffusion of the protein molecule were each obtained within the correct order of magnitude despite the short duration of the simulation. This is due to the fact that the statistics of these properties are determined by the solvent more than by the solute.

These simulations, especially that in aqueous solution, present a molecule with several interesting properties consistent with the function of parvalbumin. Thus, Molecular Dynamics simulation is a complement to experiment, even for these very complex systems. For simulation results to be more reliable in detail, though, much validation of inter- and intramolecular potentials as well as more technical aspects remains. In view of the far-reaching advances in the understanding of biomolecular mechanisms then possible, however, these issues deserve great effort.

Acknowledgement is due to Dr. Torbjörn Drakenberg for valuable discussion and encouragement, and to the Swedish Natural Science Research Council for generous allocation of computer time.

References

- (1) Frauenfelder, H., Ciba Foundation Symposium 93, "Mobility and Function in Proteins and Nucleic Acids", 329, Pitman, London, 1983.
- (2) Perutz, M. F. and Mathews, F. S., J. Mol. Biol., 21, 199 (1966).
- (3) Takano, T., J. Mol. Biol., 110, 569 (1977).
- (4) Debrunner, P. G. and Frauenfelder, H., Ann. Rev. Phys. Chem., 33, 283 (1982).

- (5) Linderström-Lang, K., *Chem. Soc. Spec. Publ.*, **2**, 1 (1955).
- (6) Hvidt, A. and Nielsen, S. O., *Adv. Prot. Chem.*, **21**, 287 (1966).
- (7) Wagner, G., *Quart. Rev. Biophys.*, **16**, 1 (1983).
- (8) Allewell, N. M., *J Biochem. Biophys. Meth.*, **7**, 345 (1983).
- (9) Englander, S. W. and Kaltenbach, N., *Quart. Rev. Biophys.*, **16**, 521 (1984).
- (10) Woodward, C. K. and Hilton, B. D., *Biophys. J.*, **32**, 561 (1980).
- (11) Praissman, M. and Rupley, J. A., *Biochemistry*, **7**, 2446 (1968).
- (12) Tüchsen, E., Hvidt, A. and Ottesen, M., *Biochemie*, **62**, 563 (1980).
- (13) Hanson, J. and Schoenborn, B., *J. Mol. Biol.*, **153**, 117 (1981).
- (14) Kossiakoff, A. A., *Nature*, **296**, 713 (1982).
- (15) Wlodauer, A. and Sjölin, L., *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 3628 (1982).
- (16) Lakowitz, J. R. and Weber, G., *Biochemistry*, **12**, 4171 (1973).
- (17) Eftink, M. R. and Ghiron, C. A., *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 3290 (1975).
- (18) Eftink, M. R. and Ghiron, C. A., *Biochemistry*, **16**, 5546 (1977).
- (19) Munro, I., Pecht, I. and Stryer, L., *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 56 (1979).
- (20) Hochstrasser, R. M. and Negus, D. K., *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 4399 (1984).
- (21) Wüthrich, K. and Wagner, G., *FEBS Lett.*, **50**, 265 (1975).
- (22) Wagner, G., DeMarco, A. and Wüthrich, K., *Biophys. Struct. & Mech.*, **2**, 139 (1976).
- (23) Gurd, F. R. N. and Rothgeb, T. M., *Adv. Prot. Chem.*, **33**, 74 (1979).
- (24) Petsko, G. A. and Ringe, D., *Ann. Rev. Biophys. Bioeng.*, **13**, 331 (1984).
- (25) McCammon, J. A., Gelin, B. R. and Karplus, M., *Nature (London)*, **267**, 585 (1977).
- (26) Gelin, B. R. and Karplus, M., *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 801 (1977).
- (27) Karplus, M. and McCammon, J. A., *Nature (London)*, **277**, 578 (1979).
- (28) McCammon, J. A., *Rep. Prog. Phys.*, **47**, 1 (1984).
- (29) Swaminathan, S. Ichiye, T., van Gunsteren, W. and Karplus, M., *Biochemistry*, **21**, 5230 (1982).
- (30) van Gunsteren, W. F., Berendsen, H. J. C., Hermans, J., Hol, W. G. J. and Postma, J. P. M., *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 4315 (1983).
- (31) Karplus, M. and McCammon, J. A., *C. R. C. Crit. Rev. Biochem.*, **9**, 293 (1981).
- (32) McCammon, J. A. and Karplus, M., *Acc. Chem. Res.*, **16**, 187 (1983).
- (33) Wnuk, W., Cox, J. A. and Stein, E. A., in "Calcium and Cell Function" (Cheung, W. Y., ed.) Vol II, 243, Acad. Press, New York, 1982.
- (34) Demaille, J. G., in "Calcium and Cell Function" (Cheung, W. Y., ed.) Vol II, 111, Acad. Press, New York, 1982.
- (35) Seamon, K. B. and Kretsinger, R. H., in "Calcium in Biology" (Spiro, T. G., ed.), 1, J. Wiley-Interscience, New York, 1983.
- (36) Kretsinger, R. H., *C.R.C. Crit. Rev. Biochem.*, **8**, 119 (1980).
- (37) Matsuoka, O., Clementi, E. and Yoshimine, M., *J. Chem. Phys.*, **64**, 1351 (1976).
- (38) Lie, G. C., Clementi, E. and Yoshimine, M., *J. Chem. Phys.*, **64**, 2314 (1976).
- (39) Clementi, E., Canallone, F. and Scordamaglia, R., *J. Am. Chem. Soc.*, **99**, 5531 (1977).
- (40) Bolis, G. and Clementi, E., *J. Am. Chem. Soc.*, **99**, 5550 (1977).
- (41) Clementi, E., Corongiu, G. and Ranghino, G., *J. Chem. Phys.*, **74**, 578 (1981).
- (42) Clementi, E., Corongiu, G., Jönsson, B. and Romano, S., *FEBS Lett.*, **100**, 313 (1979).
- (43) Lifson, S., NATO Advanced Study Institute/FEBS Advanced Course No. 78 (1981).
- (44) Levitt, M. and Lifson, S., *J. Mol. Biol.*, **46**, 269 (1969).
- (45) Hagler, A. T. and Lapicciarella, A., *Pept. Chem. Struct. Biol.*, *Proc. Am. Pept. Symp.* 4th, 279 (1975).
- (46) Hermans, J., Berendsen, H. J. C., van Gunsteren, W. F. and Postma, J. P. M., *Biopolymers*, **23**, 1513 (1984).
- (47) Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M., *J. Comp. Chem.*, **4**, 187 (1983).
- (48) van Gunsteren, W. F. and Karplus, M., *Macromolec.*, **15**, 1528 (1982).
- (49) Margenau, M. and Kestner, N. R., *Theory of Intermolecular Forces*, Pergamon Press, New York, 1969.
- (50) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. and Hermans, J., in *Intermolecular Forces*, B. Pullman (ed.), Reidel, Dordrecht, 1981.

- (51) Dolphin, D. and Wick, A. E.: "Tabulation of Infrared Spectral Data"; Wiley Interscience, New York, 1977.
- (52) Herzberg, G.: "Molecular Spectra and Molecular Structure: Infrared and Raman Spectra of Polyatomic Molecules"; van Nostrand, Princeton, 1945.
- (53) The parameters for the Cys residue were inferred from experimental data about methanethiol and ethanethiol obtained from: Schegel, H. B., Wolfe, S. and Bernardi, F., J. Chem. Phys., 67, 4181 (1977), and Durig, J. R., Bucy, W. E., Wurrey and Carreira, L. A., J. Phys. Chem., 79, 988 (1975).
- (54) Neumann, M., J. Chem. Phys., 82, 5663, (1985).
- (55) Teleman, O. and Jönsson, B., J. Comp. Chem., in press.
- (56) Gear, C. W.: "Numerical Initial Value Problems in Ordinary Differential Equations", 148-155; Prentice Hall, Englewood Cliffs, N. J., 1971.
- (57) Teleman, O., Jönsson, B. and Engström, S., to be published.
- (58) Kretsinger, R. H. and Nockolds, C. E., J. Biol. Chem., 218, 3313 (1973).
- (59) McQuarrie, D. A.: "Statistical Mechanics", p 455, Harper and Row, New York, N. Y., 1976.
- (60) Marshall, A. G.: "Biophysical Chemistry; Principles, Techniques and Applications", 712-715; John Wiley & Sons, New York, 1978.
- (61) Teleman, O. and Ahlström, P., J. Am. Chem. Soc., in press.
- (62) Suezaki, Y. and Go, N., Int. J. Peptide Protein Res., 7, 333 (1975).
- (63) See for instance C. R. Cantor & P. R. Schimmel: "Biophysical Chemistry, Part II: Techniques for the Study of Biological Structure and Function", Freeman, San Fransisco, 1980, 763-788. The procedure followed here corresponds to eq. (13-108) of that ref., but in real space and with $\kappa=1$ and W_{hkl} equal to the relative masses of the atoms.
- (64) Levitt, M., J. Mol. Biol., 168, 595 (1983).
- (65) Peticolas, W. L., Meth. Enzymol., 61, 425 (1979).
- (66) Åqvist, J., van Gunsteren, W. F., Leijonmarck, M. and Tapia, O., J. Mol. Biol., 183, 461 (1985).
- (67) Reimers, J. R. and Watts, R. O., Chem. Physics, 91, 201 (1984).
- (68) Probst, M. M., Radnai, T., Heinzinger, K., Bopp, P. and Rode, B. M., J. Phys. Chem. **1985**, 89, 753-759.
- (69) Teleman, O., Drakenberg, T., Forsén, S. and Thulin, E., Eur. J. Biochem., 134, 453 (1983).

Tables

Table I. Simulation parameters and averages.

	APO	VAC	AQ
Number of atoms	983	985	7970
Box size (Å)	∞	∞	43x50x43
Time steps $\Delta t/\Delta T$ (fs)	0.2/1.0	0.2/1.0	0.2/1.2
Equilibration time (ps)	29.	29.	36.
Length of trajectory (ps)	80.	80.	70.
Velocity scaling interval (fs)	200	200	99.6
Neighbour list update interval (fs)	12	12	4.8
CPU time/ps (s) ^{a),b)}	479	479	2523
CPU time/pair interaction (μs) ^{a)}	5.7	5.7	2.9
Average temperature (K)	305	306	303
Average pressure (MPa)	-	-	143

^{a)}The difference in timing between the *in vacuo* and solution simulations partly reflects improvements in the MD program and in the compiler. ^{b)} All simulations were performed on a Cray-1A and timing refers to total CPU time.

Table II. R factors as function of time. The R factors (see Eq. 14) for the backbone atom positions of parvalbumin compared to the X-ray coordinates from each of the three simulations.

Time after start of the simulation (ps)	APO	VAC	Time after start of the simulation (ps)	AQ
29	3.05	2.93		
39	3.14	3.01	36	2.16
49	3.21	3.19	46	2.37
59	3.22	3.33	56	2.52
69	3.14	3.36	66	2.43
79	3.27	3.35	76	2.57
89	3.19	3.37	86	2.44
99	3.35	3.46	96	2.53
109	3.32	3.59	106	2.72
Mean	3.21	3.29		2.47
R M S deviation	0.09	0.20		0.15

Table III. Mean distances in Å between the centroids of the eight phenylalanines in the core.

Residue	Phe 24	Phe 29	Phe 30	Phe 47	Phe 66	Phe 70	Phe 85
Phe 29							
X-RAY	5.83						
APO	3.88						
VAC	4.14						
AQ	5.18						
Phe 30							
X-RAY	7.75	6.42					
APO	5.17	5.23					
VAC	6.61	6.40					
AQ	5.40	6.11					
Phe 47							
X-RAY	16.79	16.23	12.40				
APO	13.06	11.95	10.86				
VAC	19.12	18.86	14.04				
AQ	12.22	15.30	9.58				
Phe 66							
X-RAY	6.31	5.70	6.41	11.57			
APO	8.24	6.85	7.81	5.94			
VAC	7.85	8.91	5.22	11.82			
AQ	6.81	9.36	5.26	7.12			
Phe 70							
X-RAY	9.44	4.92	8.10	13.80	5.22		
APO	11.29	8.39	10.41	7.58	5.10		
VAC	6.72	6.05	5.21	13.37	4.46		
AQ	8.63	8.30	6.09	10.95	5.10		
Phe 85							
X-RAY	5.63	8.75	6.52	12.02	5.36	10.28	
APO	5.93	6.20	5.56	7.42	4.06	8.21	
VAC	5.87	8.06	4.93	16.18	5.34	7.52	
AQ	4.48	9.26	6.59	9.16	5.64	9.76	
Phe 102							
X-RAY	11.78	10.97	6.34	6.08	7.36	9.80	7.62
APO	9.80	8.75	6.18	5.25	5.70	7.84	5.49
VAC	13.64	13.42	7.50	7.81	7.73	9.48	10.19
AQ	8.46	11.40	5.75	4.53	5.44	9.22	6.16

Table IV. Distribution of hydrogen bond life times in the protein molecule. For each simulation two formation distances (see text) have been used.

Simulation	APO	APO	VAC	VAC	AQ	AQ
Formation distance ($d_s/\text{\AA}$)	2.1	2.3	2.1	2.3	2.1	2.3
Life time (ps)						
< 10	413	3137	399	2731	537	3650
10 - 20	63	186	45	168	63	123
20 - 30	22	78	28	41	19	25
30 - 40	7	24	11	29	4	11
40 - 50	14	11	12	18	6	6
50 - 60	3	11	2	12	2	3
> 60	3	9	8	15	1	2

Table V. Calcium ligands in the VAC and the AQ simulations. The times in picoseconds after start of the analysis during which ligandation occurs are given in parantheses. (eq=equilibration period, o=omitted in the simulation, *=distance oscillating between 2.3 and 4.0 Å in the period 44-58 ps).

X-ray ligands bound to the calcium ions during the whole simulation

<u>CD site</u>		<u>EF site</u>	
VAC	AQ	VAC	AQ
O _{δ1} Asp 51	O _{δ1} Asp 51	O _{δ2} Asp 90	O _{δ2} Asp 90
O _{δ1} Asp 53	-	O _{δ1} Asp 92	O _{δ1} Asp 92
O _γ Ser 55	-	O _{δ2} Asp 92	-
O _{pep} Phe 57	-	O _{δ1} Asp 94	O _{δ1} Asp 94
O _{ε1} Glu 59	O _{ε1} Glu 59	O _{pep} Lys 96	-
O _{ε2} Glu 62	O _{ε2} Glu 62	O _{ε1} Glu 101	-
	O _{ε2} Glu 101	O _{ε2} Glu 101	

X-ray ligands bound to the calcium ions during a part of the analysis

<u>CD site</u>		<u>EF site</u>	
VAC	AQ	VAC	AQ
	O _{pep} Phe 57 (-eq,8-)		O _{δ2} Asp 92 (-eq,12-)
	O _{δ1} Asp 53 (-2)		O _{pep} Lys 96 (-59)
			O _{ε1} Glu 101 (-24,44-)
			O water 12 (-10)

X-ray ligands not bound to the calcium ions at all during the analysis

<u>CD site</u>		<u>EF site</u>	
VAC	AQ	VAC	AQ
	O _γ Ser 55(-eq)	O water 12(o)	

(contd.)

Table V. (cont.)

 Other ligands to the calcium ions

<u>CD site</u>		<u>EF site</u>	
VAC	AQ	VAC	AQ
O _{δ2} Asp 53 (eq-)	O _{δ2} Asp 53 (eq-)	O _{δ1} Asp 90 (7-)	O _{δ1} Asp 90 (eq-1)
	O _{ε2} Glu 59 (eq-)		O _{δ2} Asp 94 (24-44,*,58-)
O _{ε1} Glu 62 (eq-)	O _{ε1} Glu 62 (eq-)		O water 1973 (3-)
	O water 1987 (eq-)		

Table VI. Estimated correlation times in ns for residues in the calcium binding sites.

Residue	Atoms	APO	VAC	AQ
<u>CD site</u>				
Asp 51	$C_{\beta}-C_{\gamma}$	1.9	1.7	0.8
Asp 53	$C_{\beta}-C_{\gamma}$	1.2	1.9	0.3
Glu 59	$C_{\beta}-C_{\delta}$	0.1	>4.0	0.5
Glu 62	$C_{\beta}-C_{\delta}$	1.3	0.2	0.4
<u>EF site</u>				
Asp 90	$C_{\beta}-C_{\gamma}$	0.1	0.8	2.3
Asp 92	$C_{\beta}-C_{\gamma}$	0.3	2.4	0.3
Asp 94	$C_{\beta}-C_{\gamma}$	0.1	3.2	0.2

Figure captions

Fig. 1. Schematic drawing of the X-ray structure of the backbone of carp parvalbumin. The helices are named A-F according to standard notation. On top the two calcium binding sites.

Fig. 2. Plot of mean values for backbone dihedral angles Ψ and Φ a) in the X-ray structure, b) in the APO simulation, c) in the VAC simulation and d) in the AQ simulation.

Fig. 3. Standard deviations of Ψ (dashed line) and Φ (solid line) along the chain a) in the APO simulation, b) in the VAC simulation and c) in the AQ simulation.

Fig. 4. Correlation times for all C_{α} - C_{α} vectors along the protein backbone a) in the APO simulation, b) in the VAC simulation and c) in the AQ simulation.

Fig. 5. Radial distribution of protein oxygens around the calcium ions for a) the CD site, and b) the EF site. Solid lines refer to the VAC simulation and dashed lines to the AQ simulation.

Fig. 6. Estimated correlation times (τ_1) for the vector between the two outermost non-hydrogen atoms in all side chains ordered after residue type a) in the APO simulation, b) in the VAC simulation and c) in the AQ simulation.

Fig. 7. Time correlation functions for vectors in some residues. Solid lines refer to the VAC, dashed lines to the AQ and dot-dashed lines to the APO simulation. a) $O_{\delta 1}-O_{\delta 2}$ in Asp 8. This residue is situated at the surface of the protein. b) $C_{\beta}-C_{\gamma 1}$ in Val 43 which is situated at the end of the C-helix opposite the CD site. c) $C_{\gamma 2}-C_{\delta 1}$ in Ile 50. It is directed into the core and is situated near the CD site. d) $C_{\beta}-C_{\gamma}$ in Asp 94, a calcium ligand in the EF site. e) $O_{\epsilon 1}-O_{\epsilon 2}$ in Glu 60, which is situated in the CD site, but does not ligand the calcium ion.

Fig. 8. Sum over all phenylalanines of time correlation functions ($C_1(t)$) for **a)** GZ vectors and **b)** DD vectors. Solid lines refer to the VAC, dashed lines to the AQ and dot-dashed lines to the APO simulation.

Fig. 9. Time correlation functions ($C_1(t)$) for vectors in Phe 30 for **a)** GZ vectors and **b)** DD vectors. Solid lines refer to the VAC, dashed lines to the AQ and dot-dashed lines to the APO simulation..

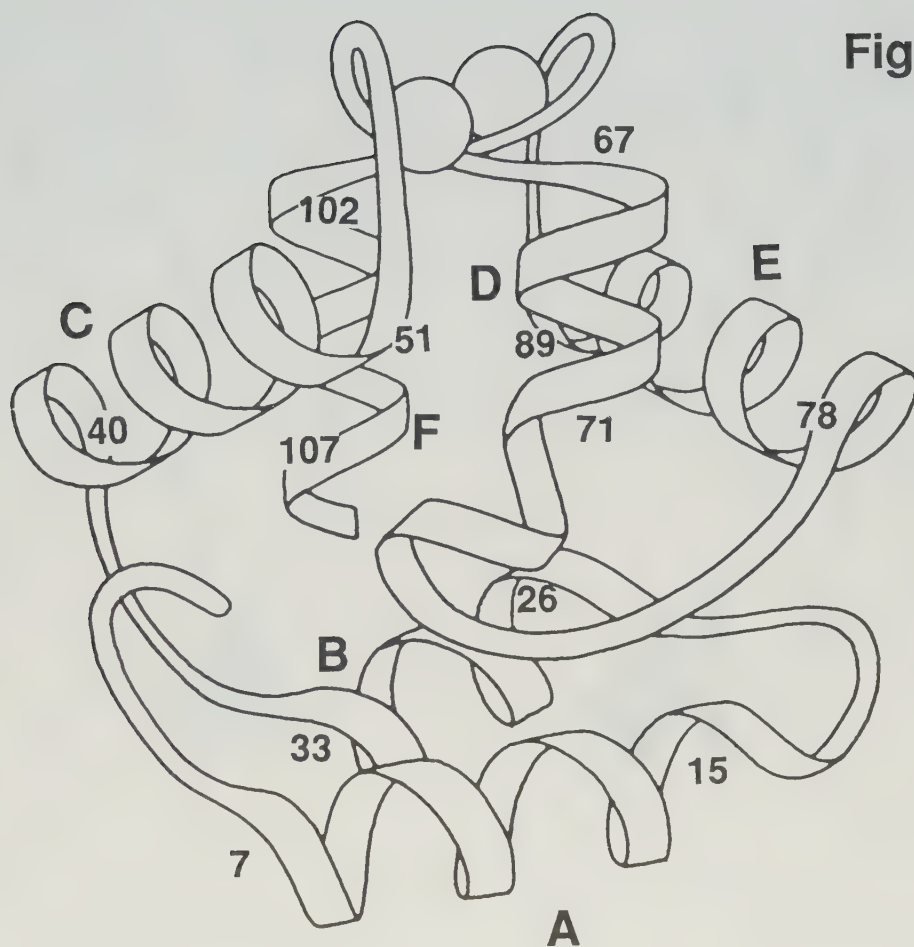


Fig. 1

Carp Muscle Calcium-binding Protein

Fig. 2a

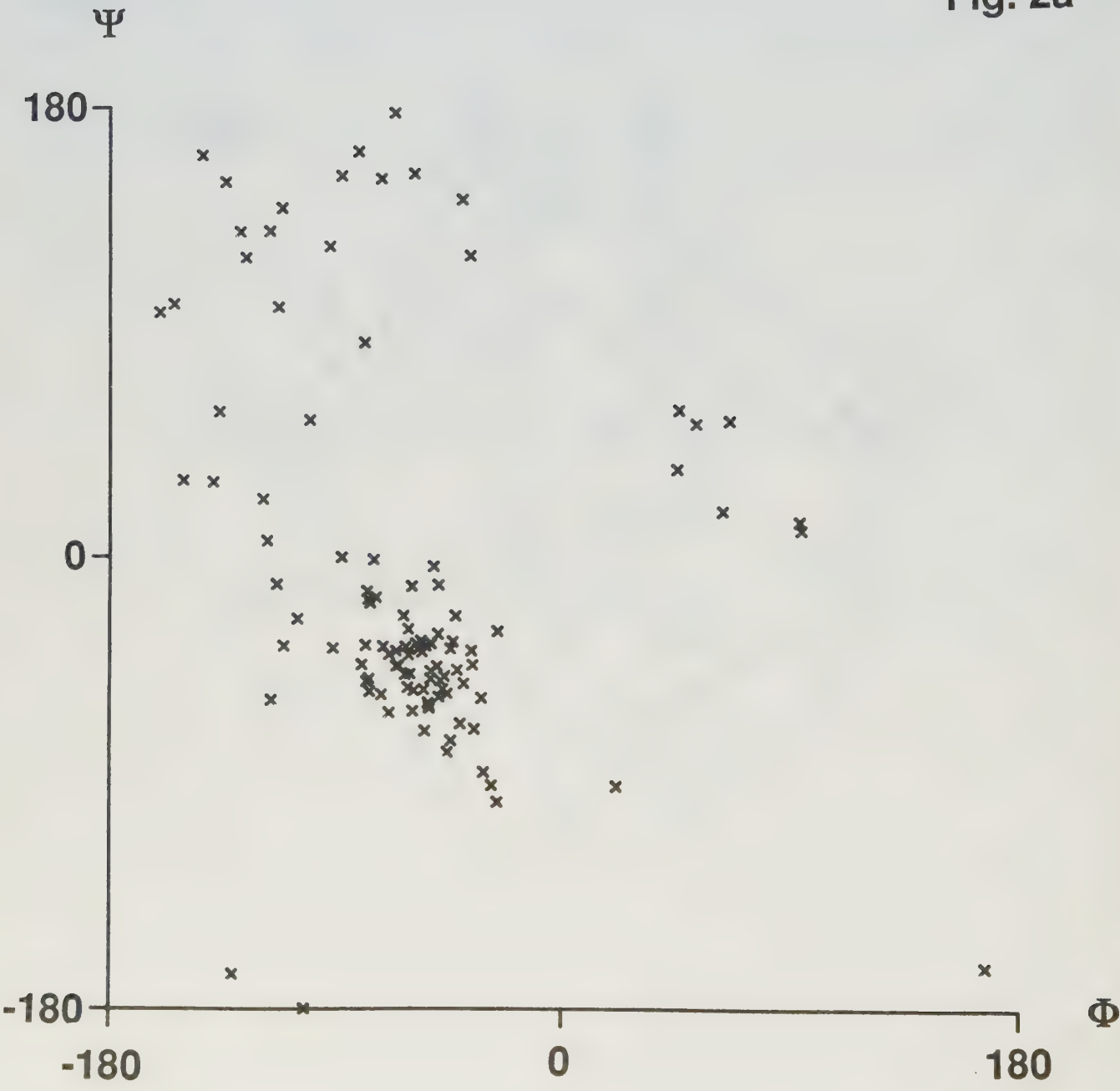


Fig. 2b

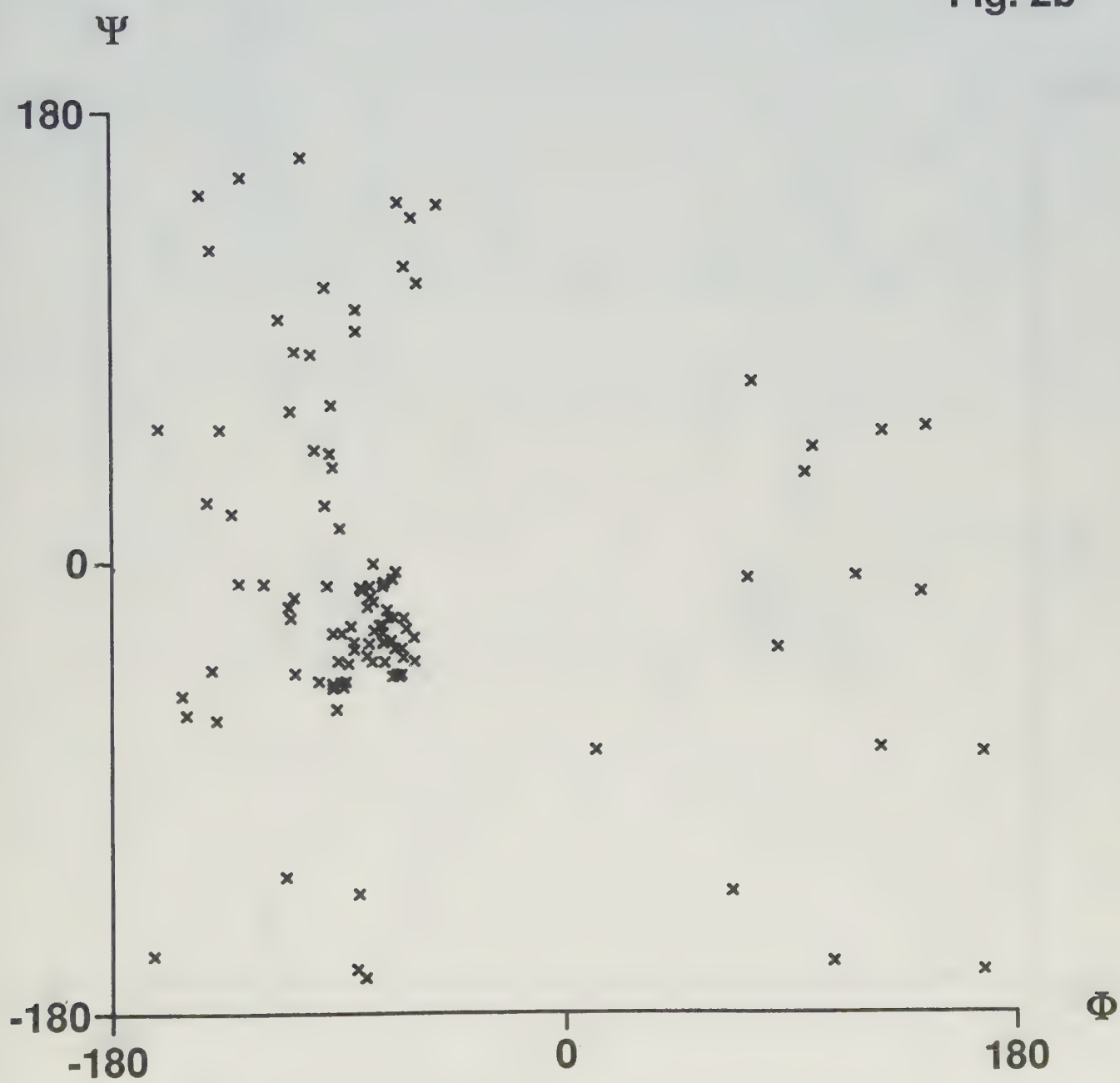


Fig. 2c

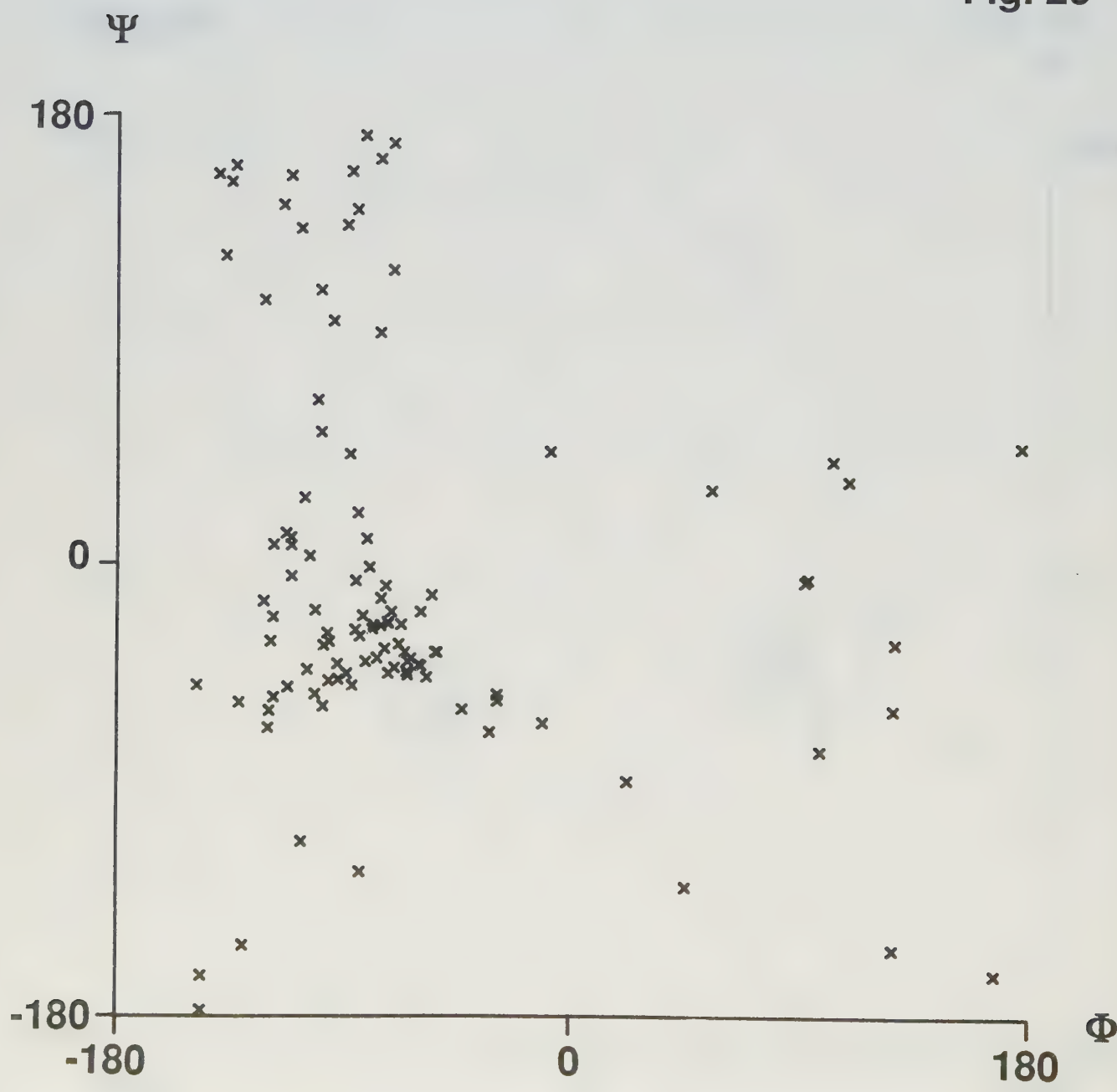
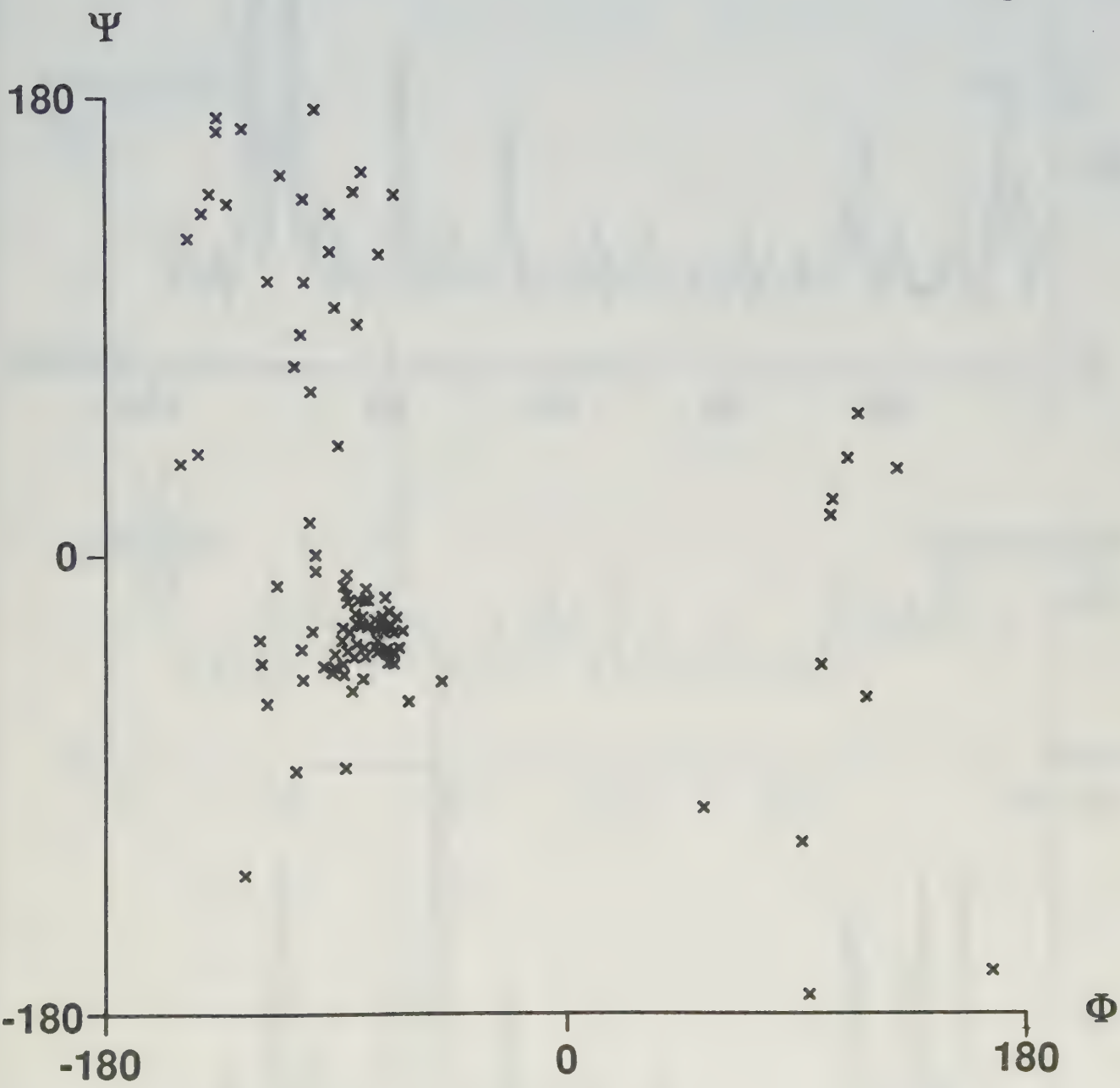
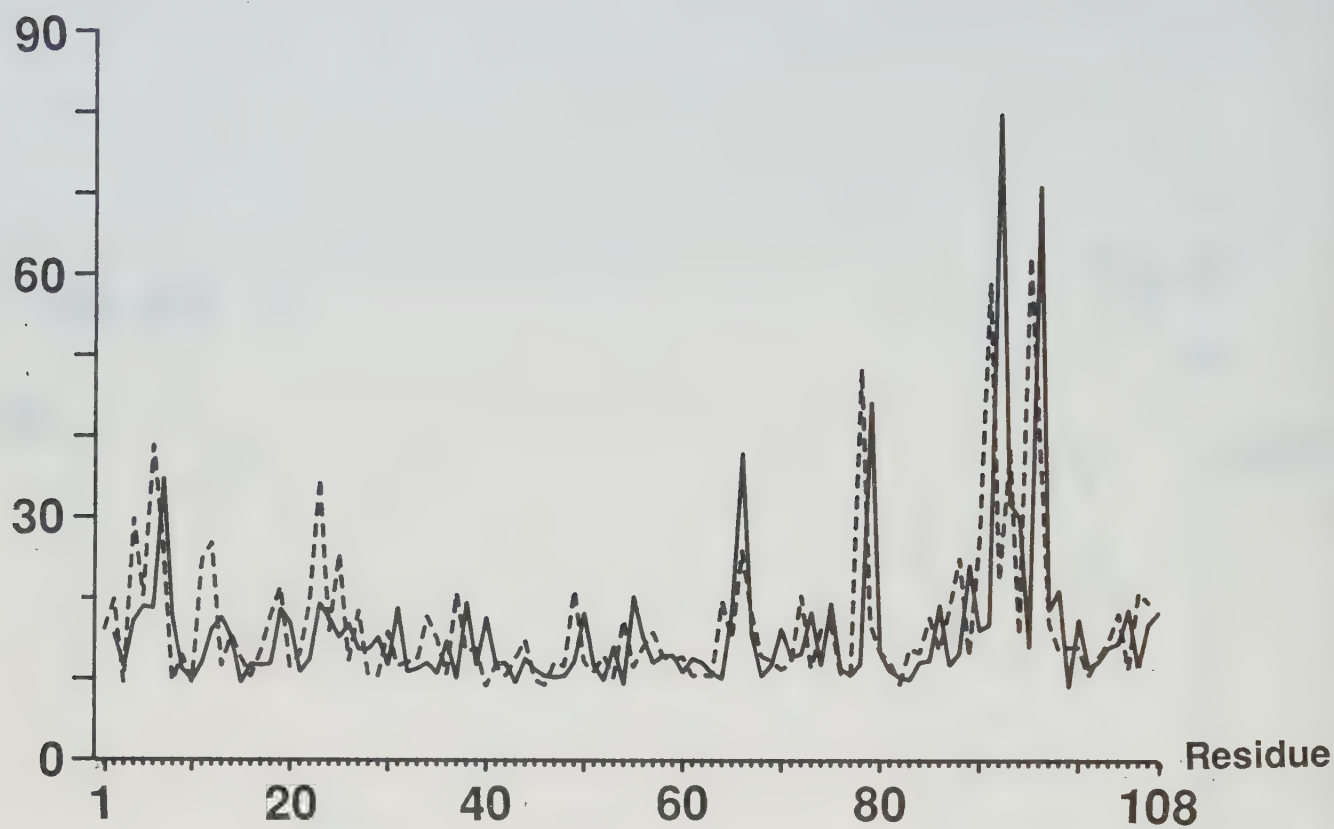


Fig. 2d



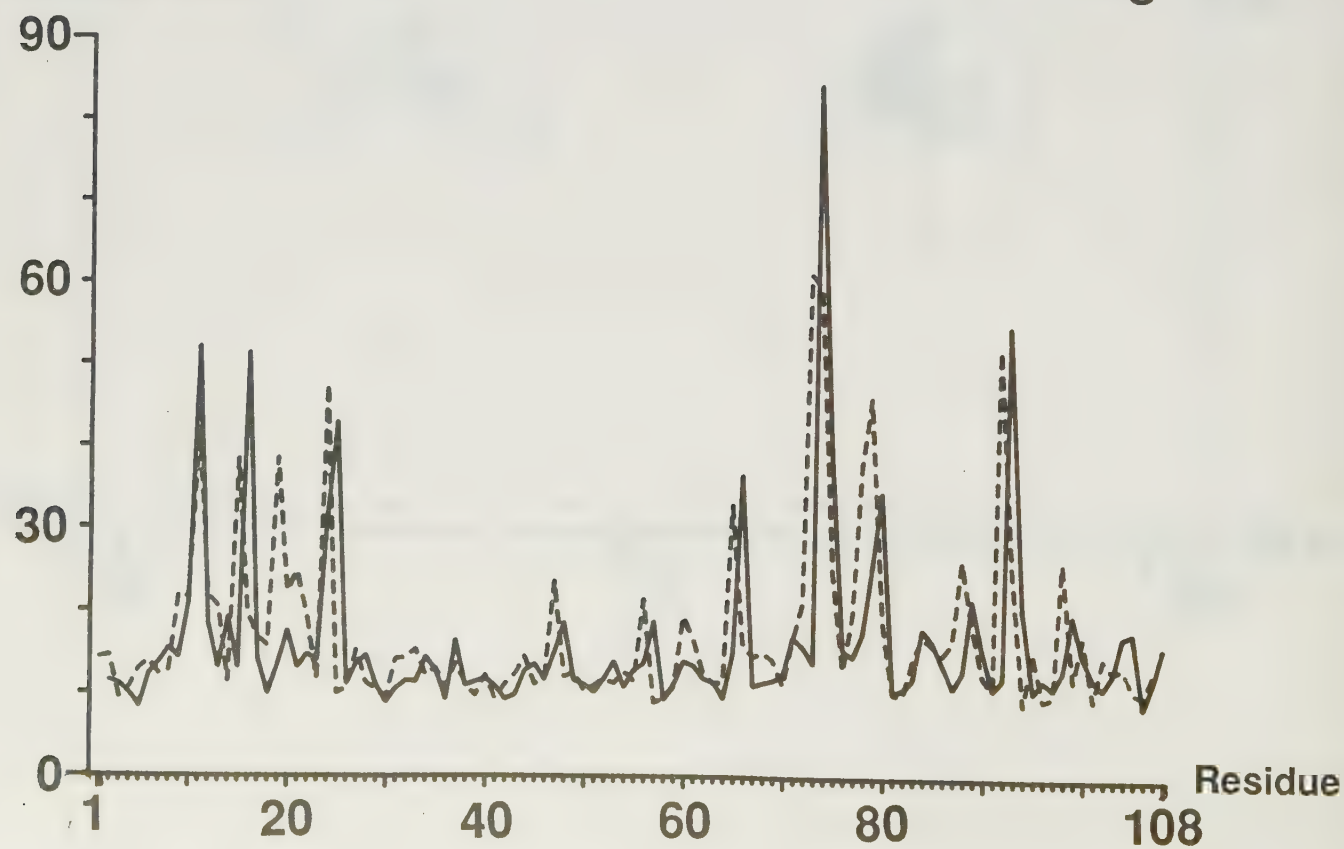
$\sigma(\text{degrees})$

Fig. 3a



$\sigma(\text{degrees})$

Fig. 3b



$\sigma(\text{degrees})$

Fig. 3c

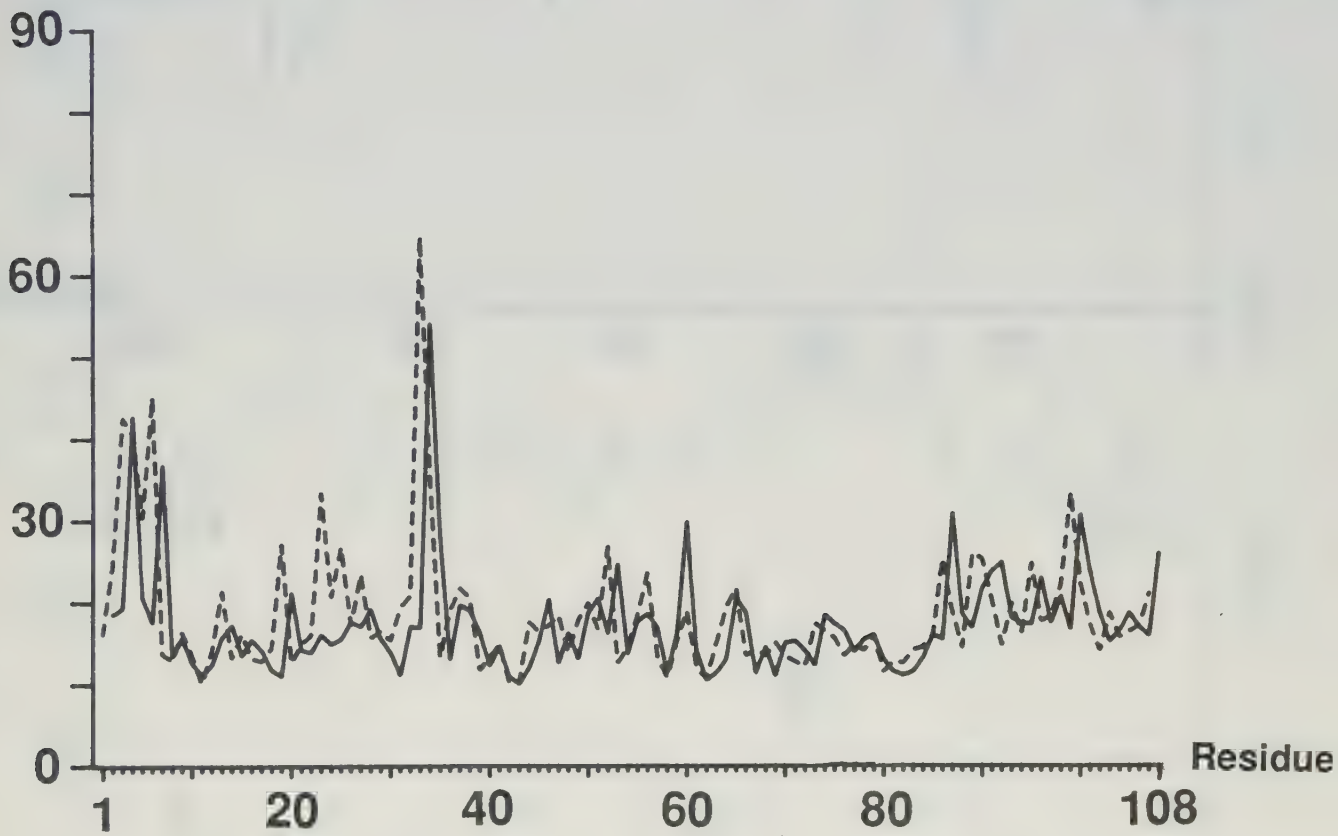


Fig. 4a

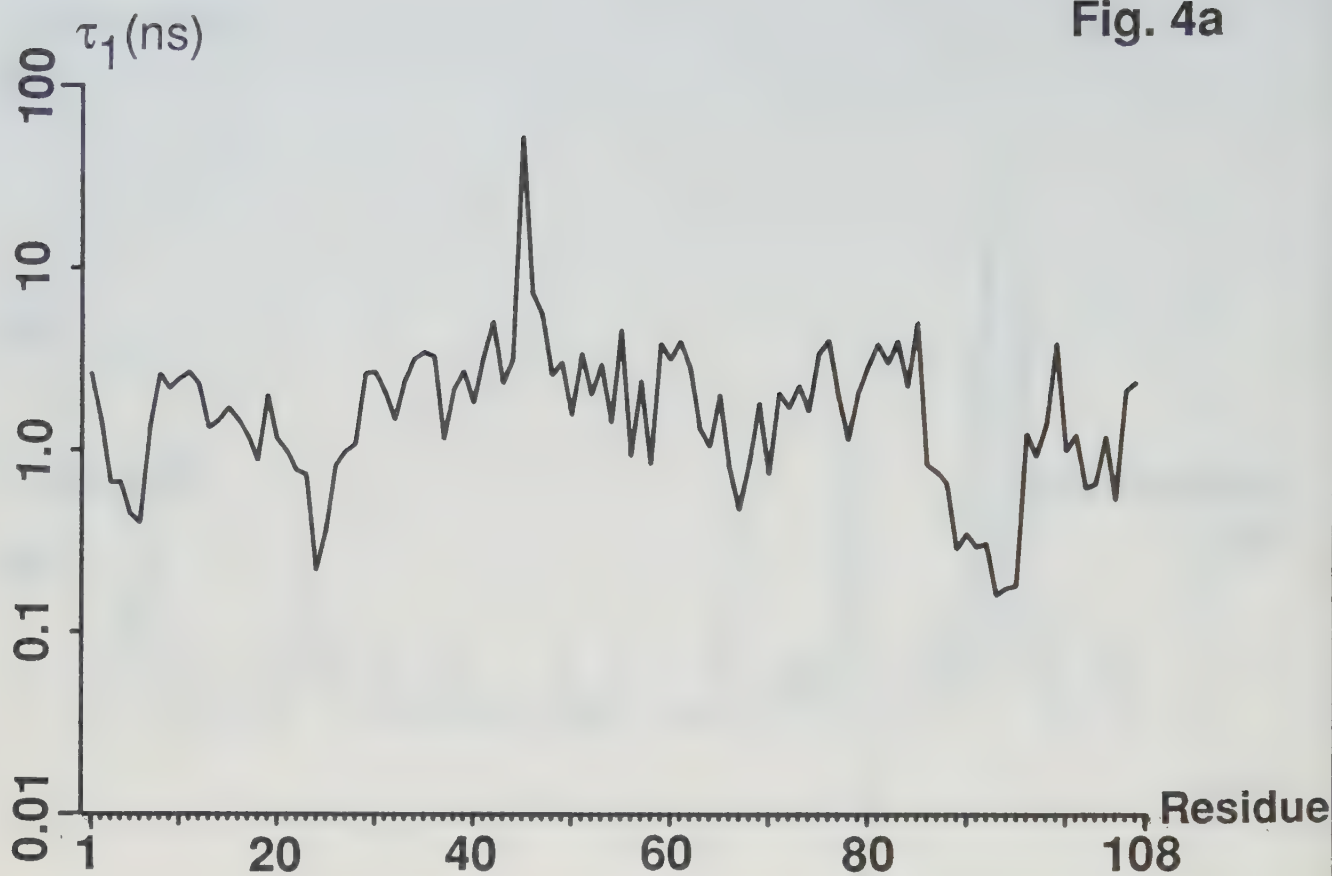


Fig. 4b

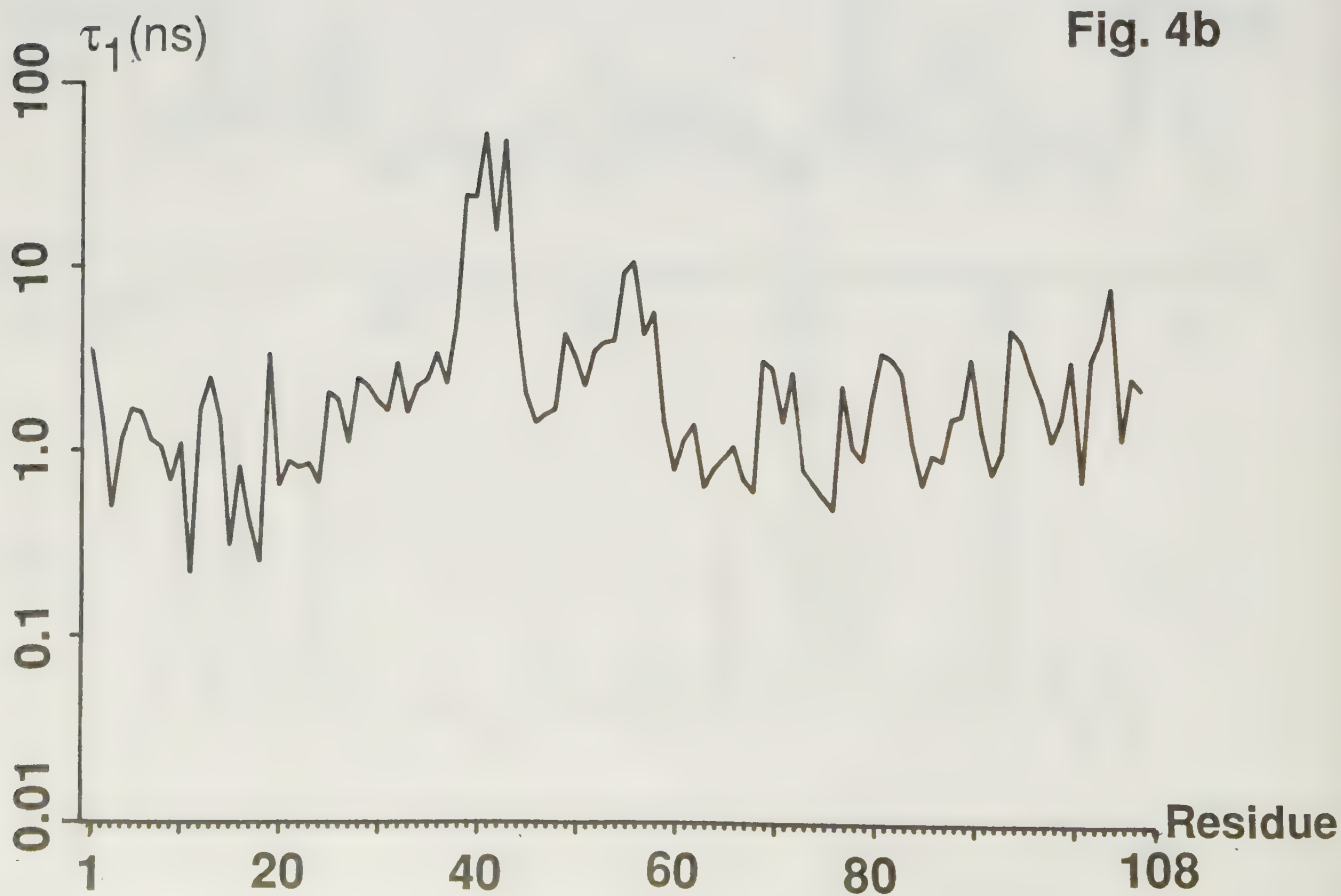


Fig. 4c

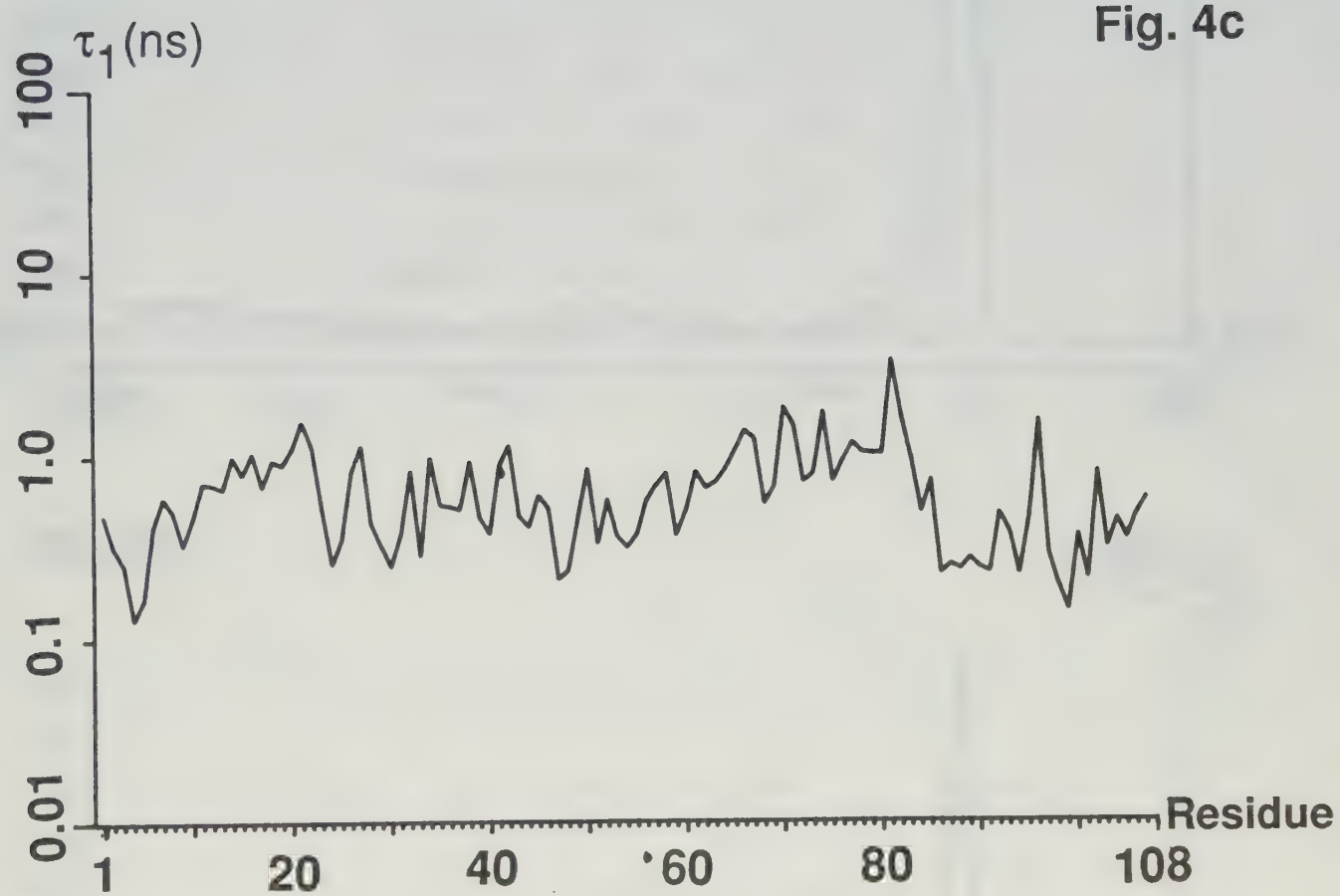


Fig. 5a

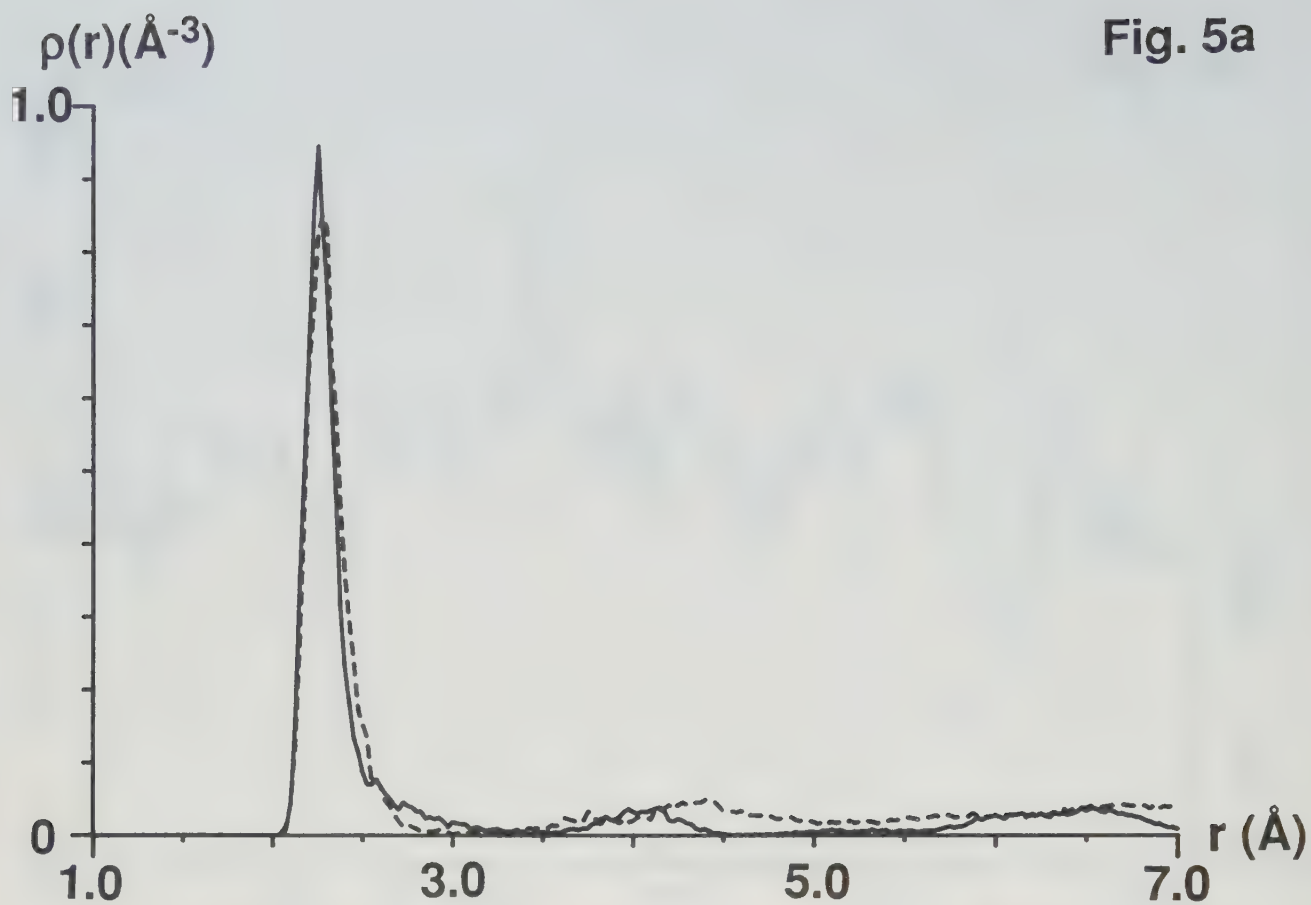


Fig. 5b

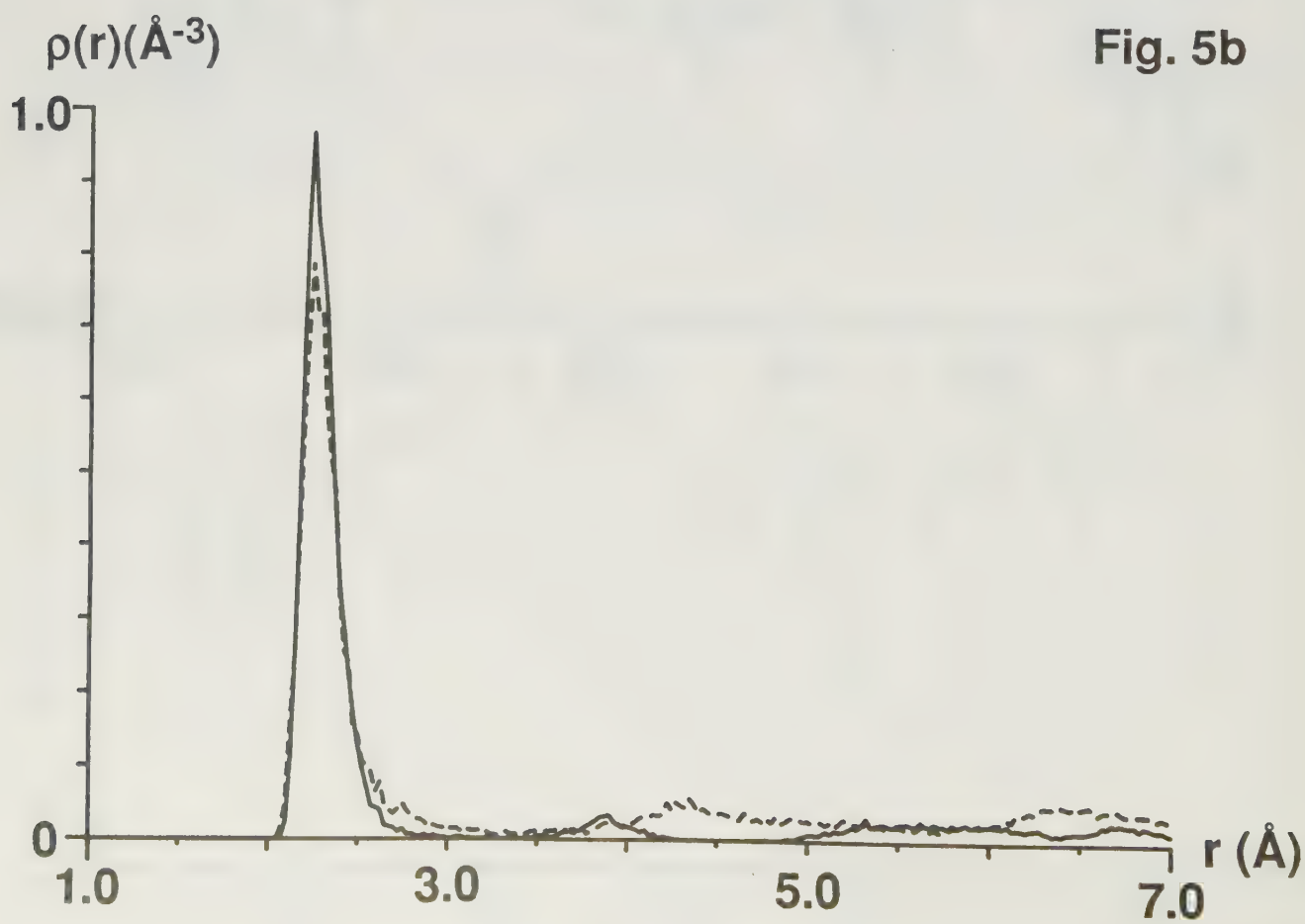


Fig. 6a

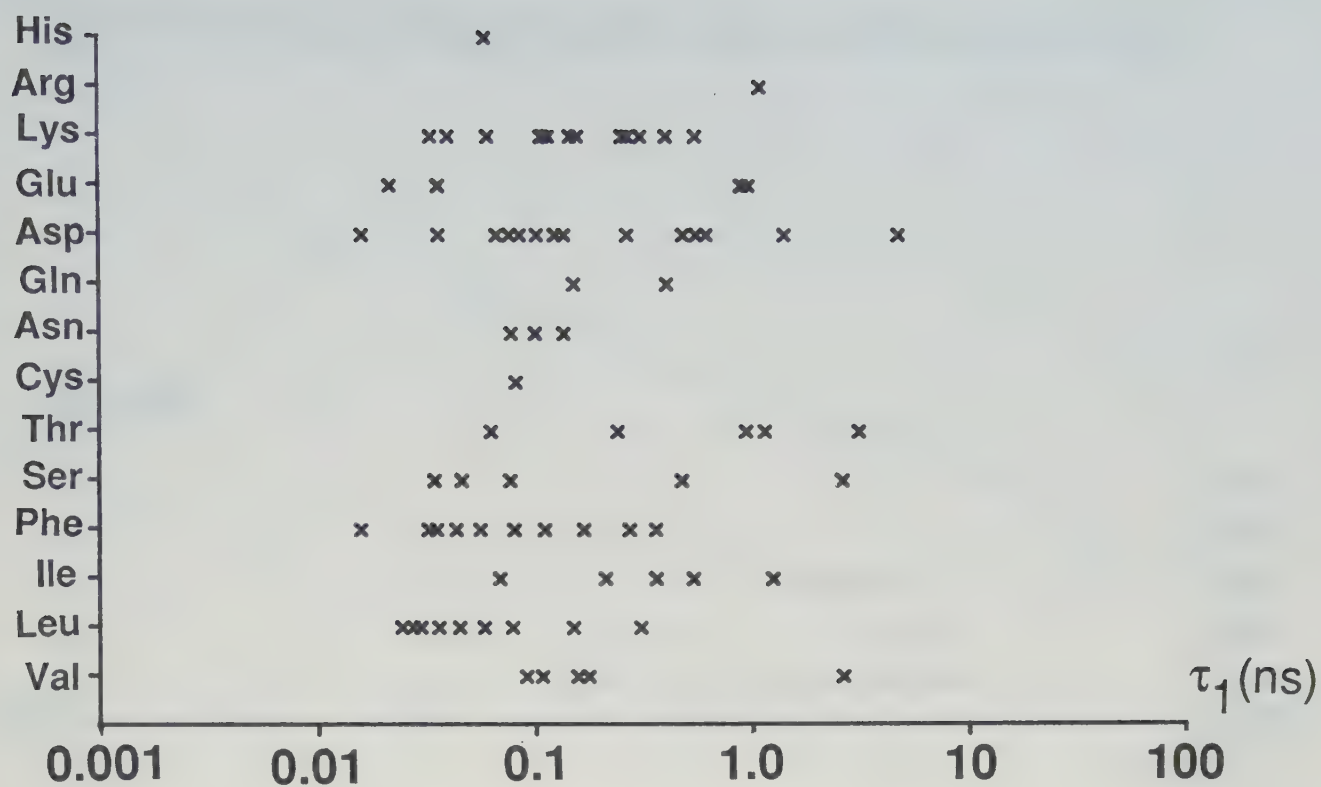


Fig. 6b

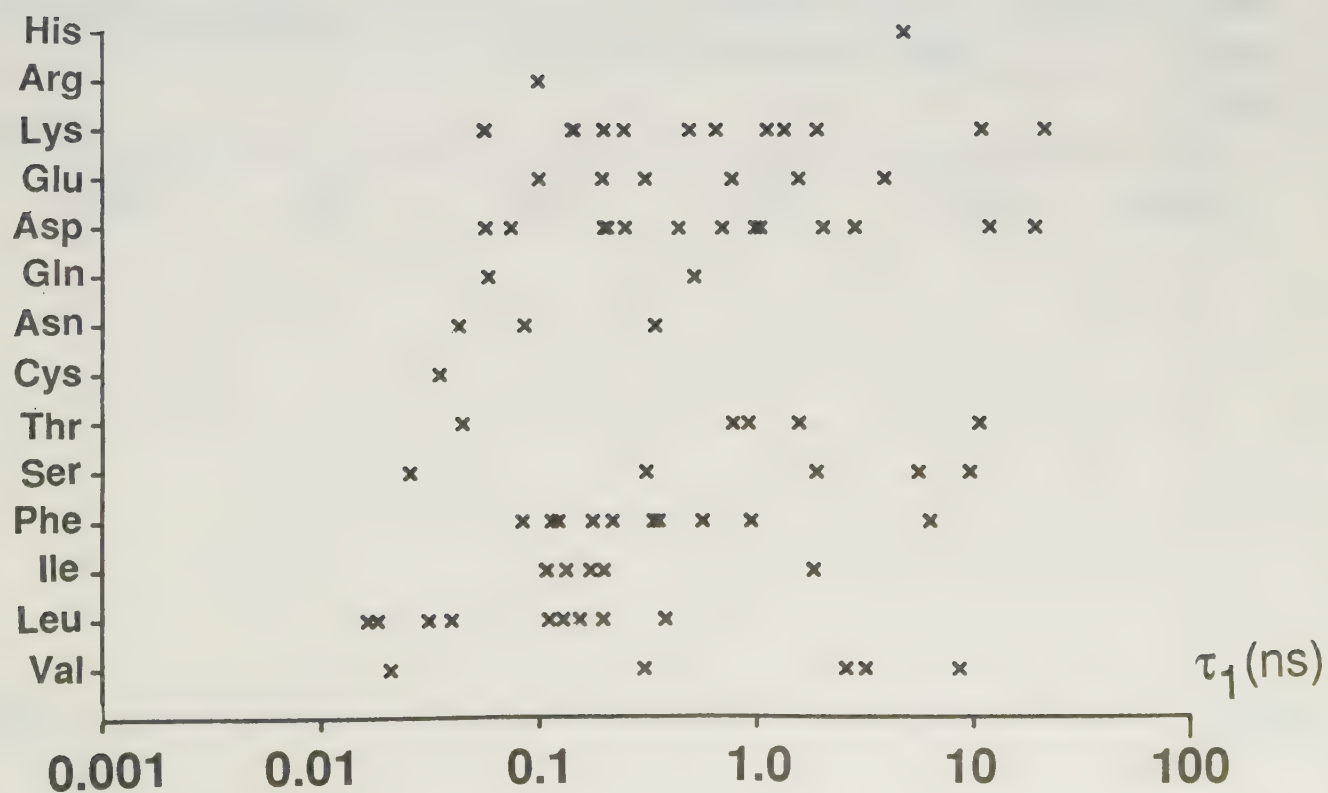


Fig. 6c

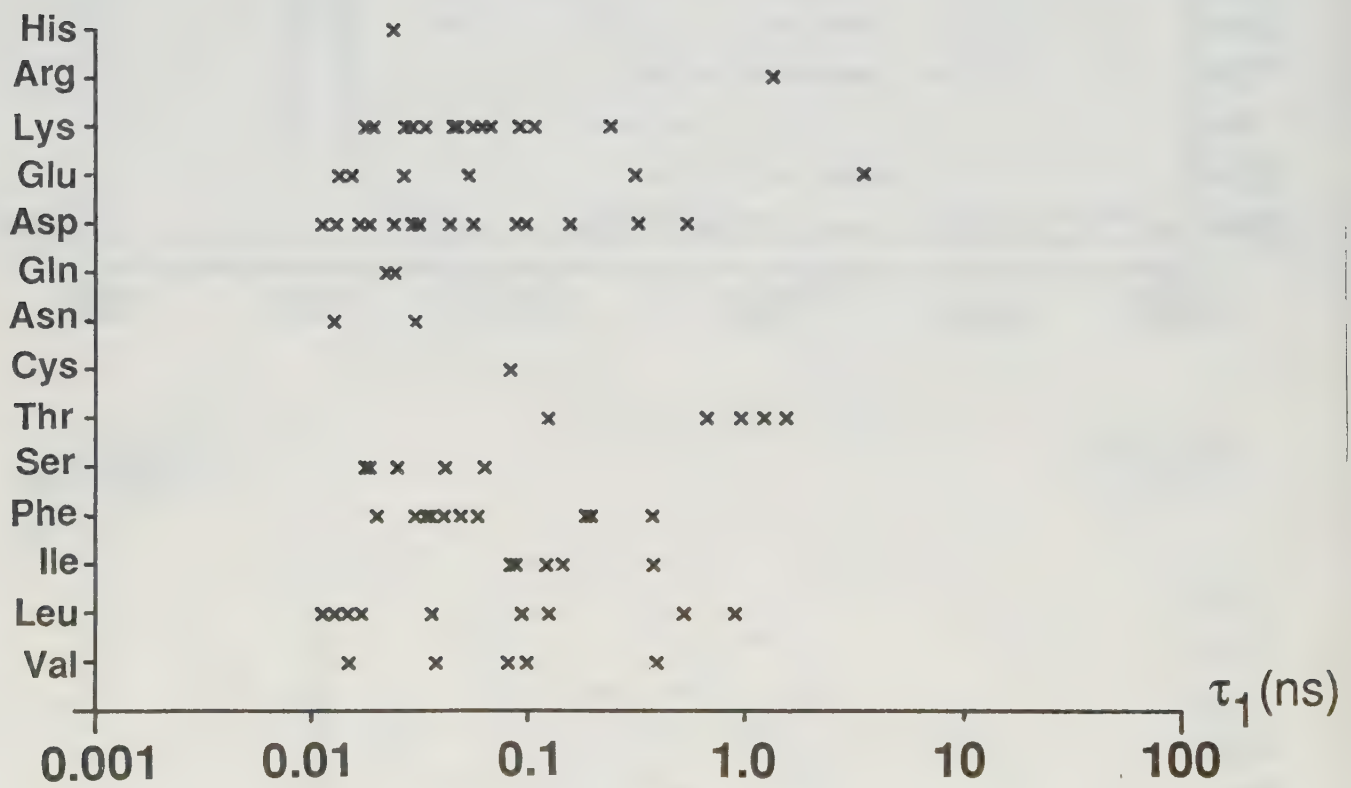


Fig. 7a

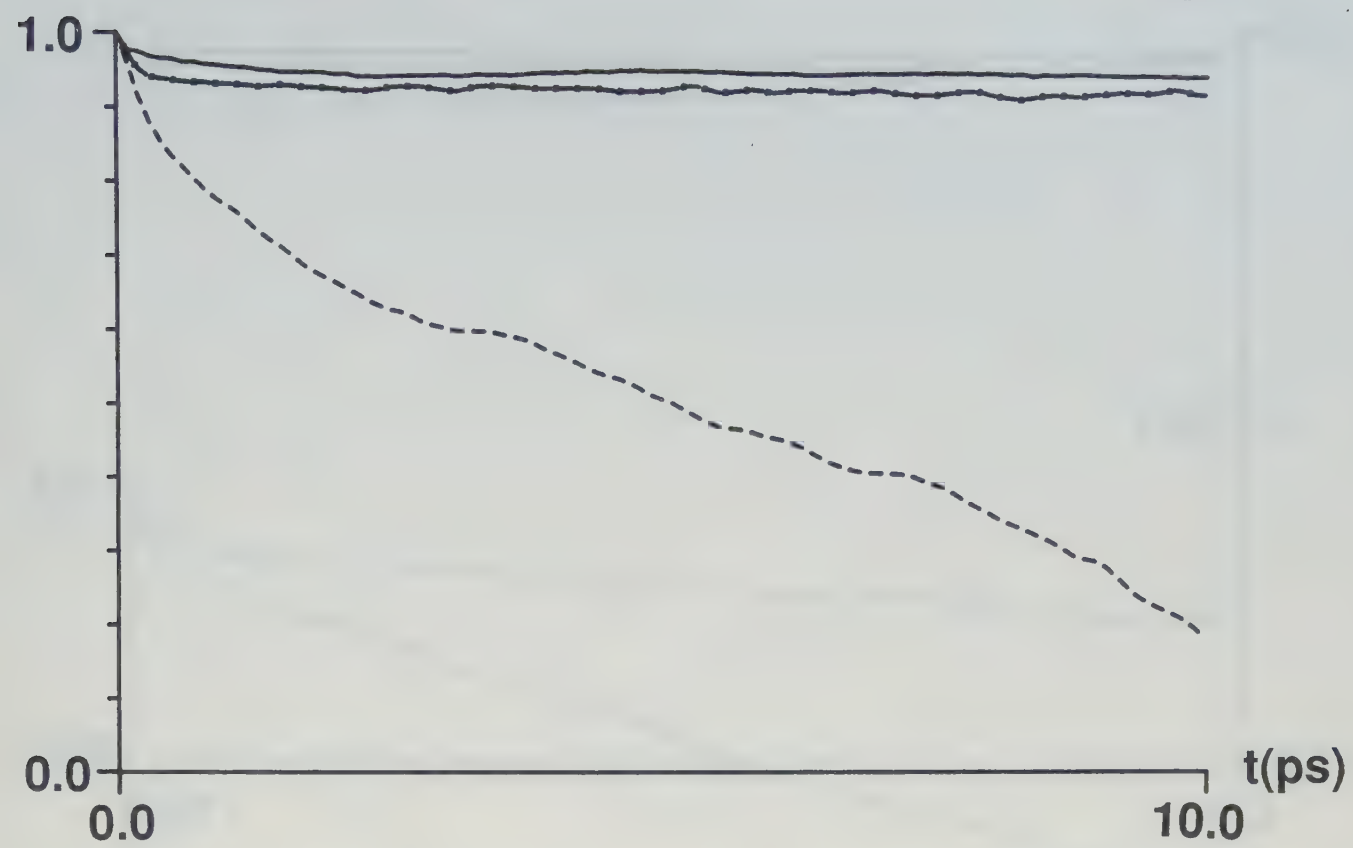


Fig. 7b

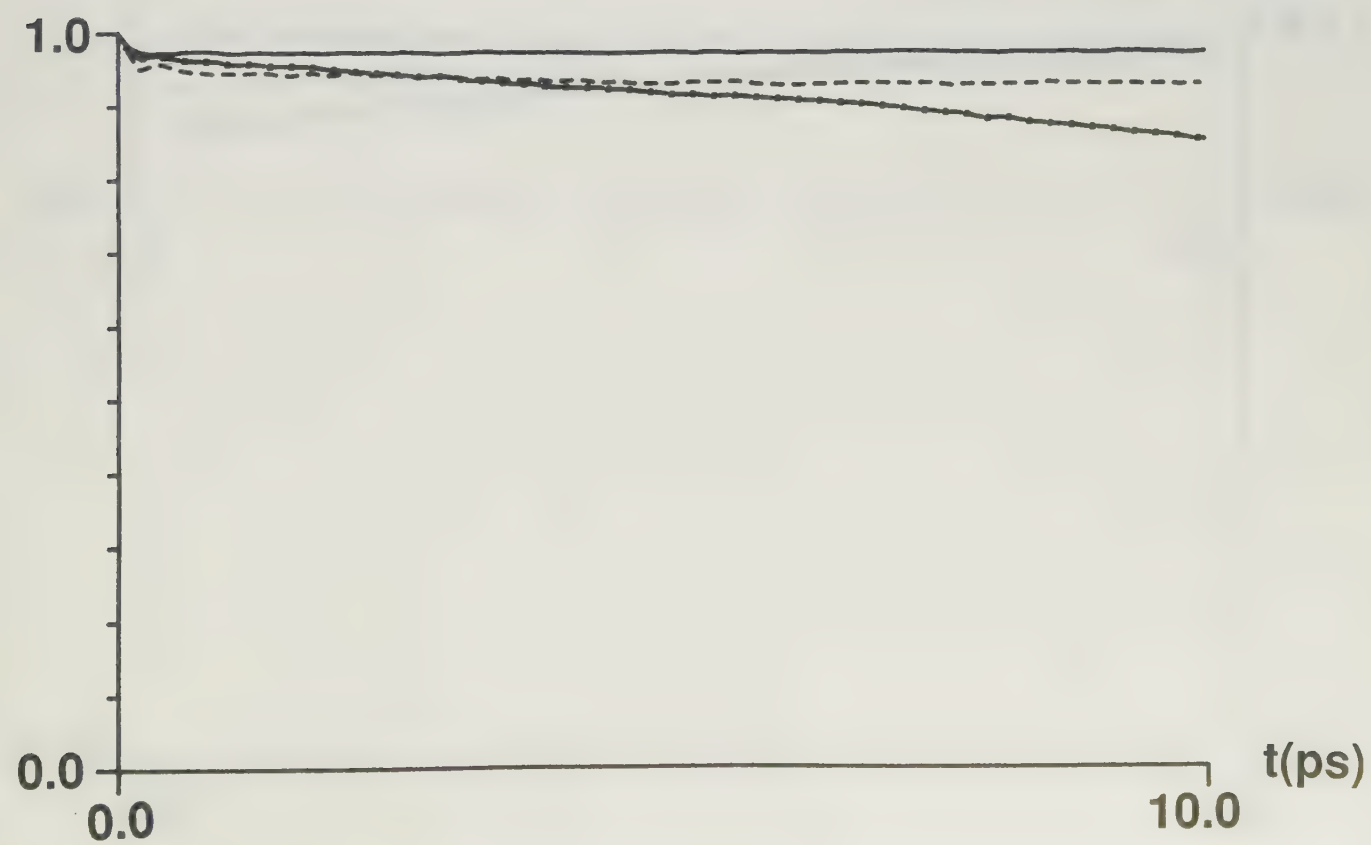


Fig. 7c

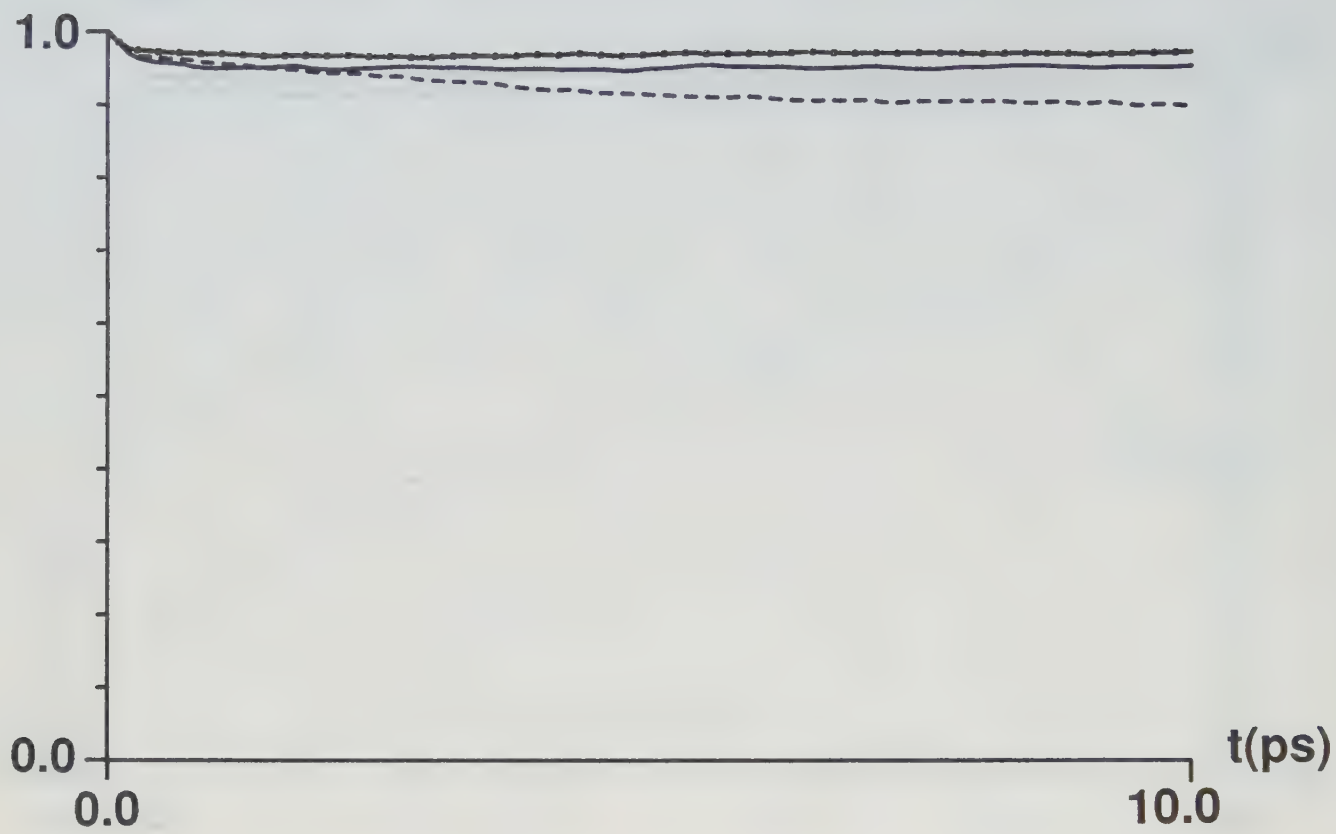


Fig. 7d

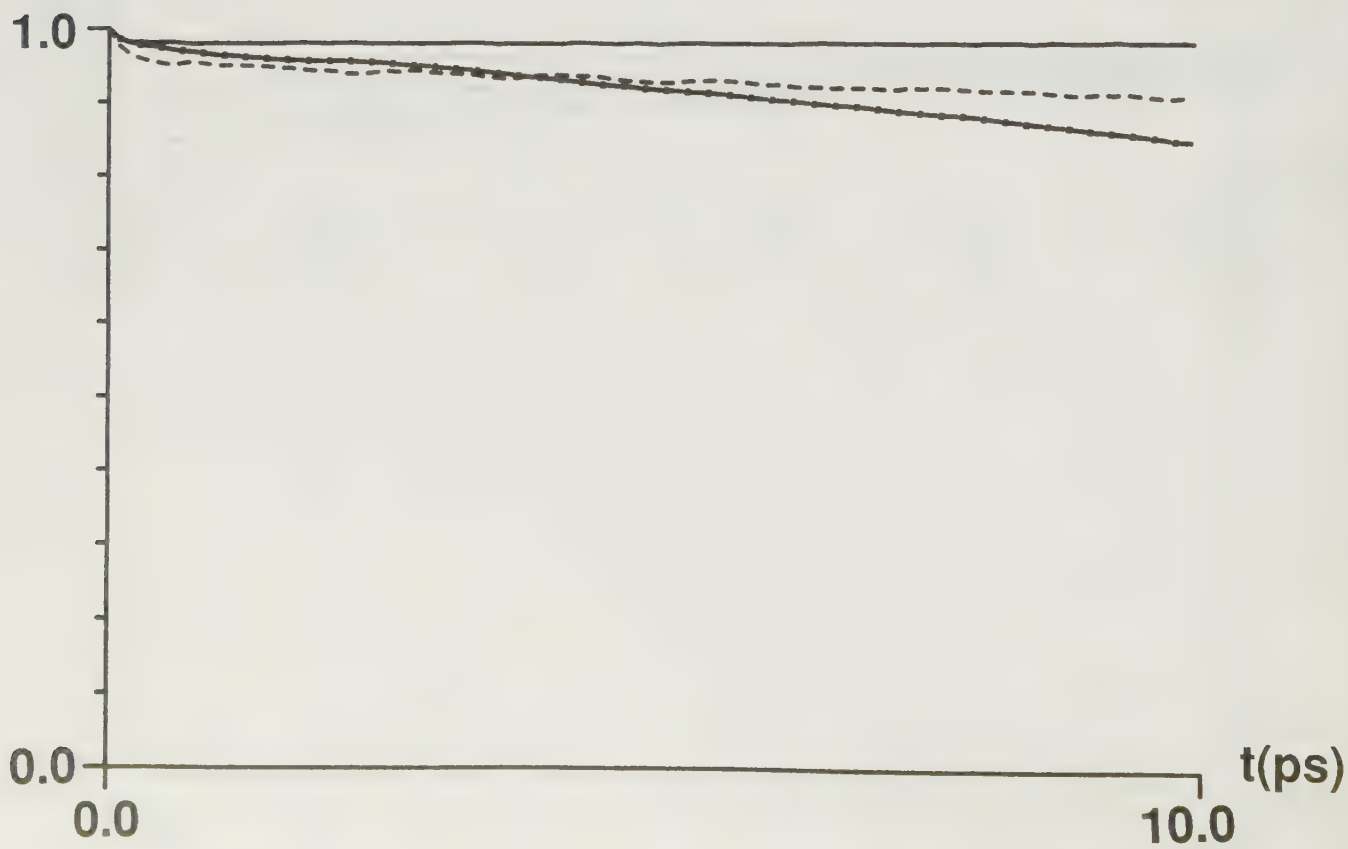


Fig. 7e

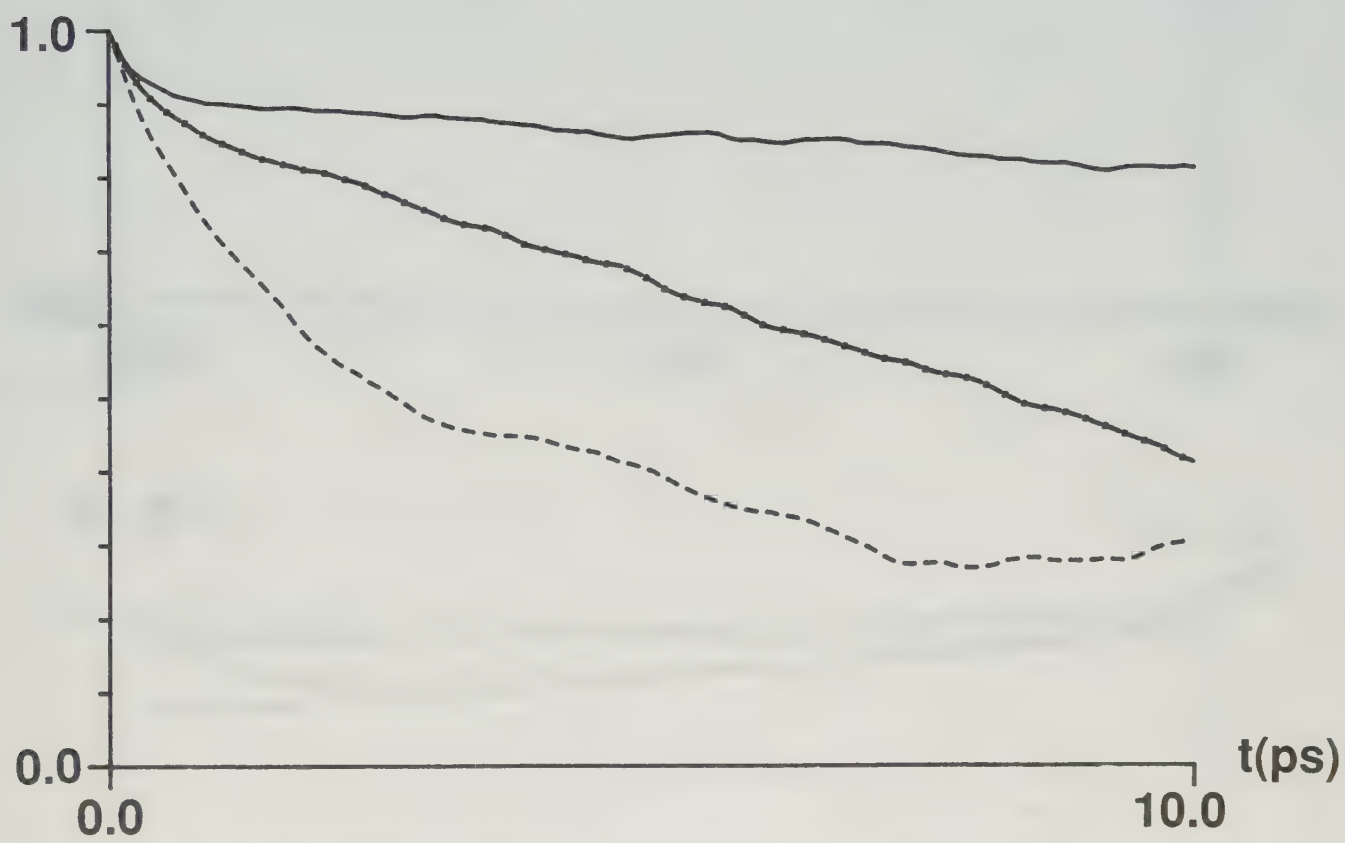


Fig. 8a

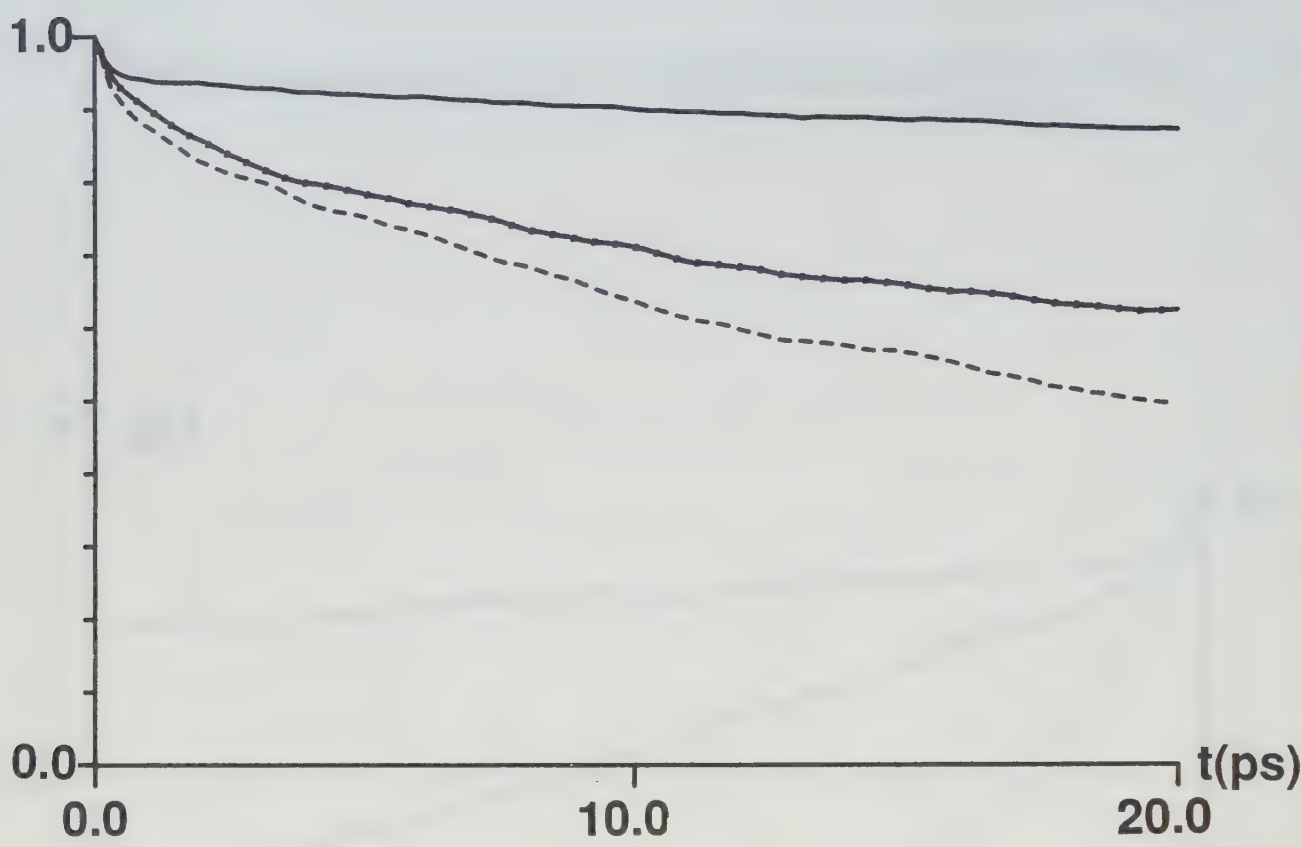


Fig. 8b

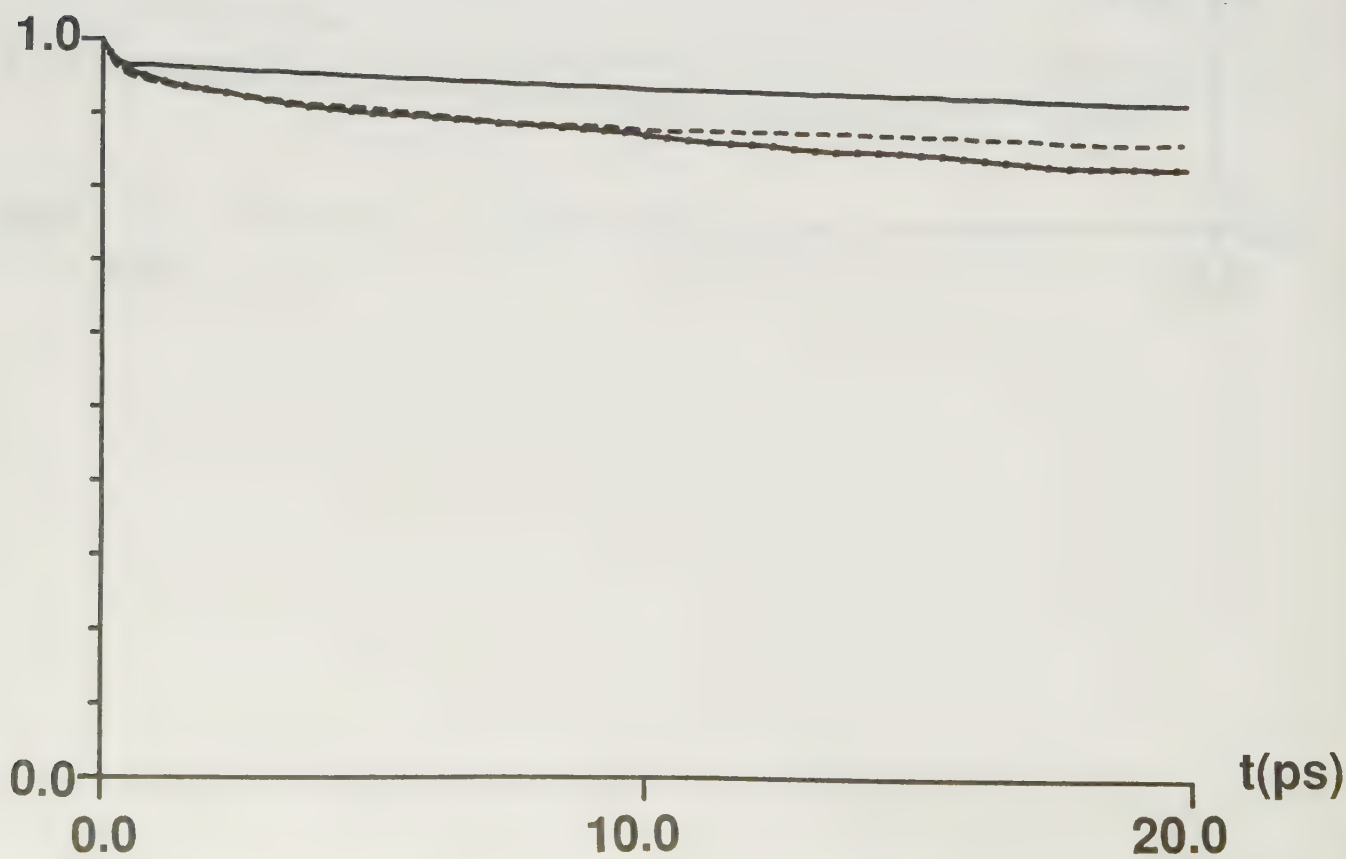


Fig. 9a

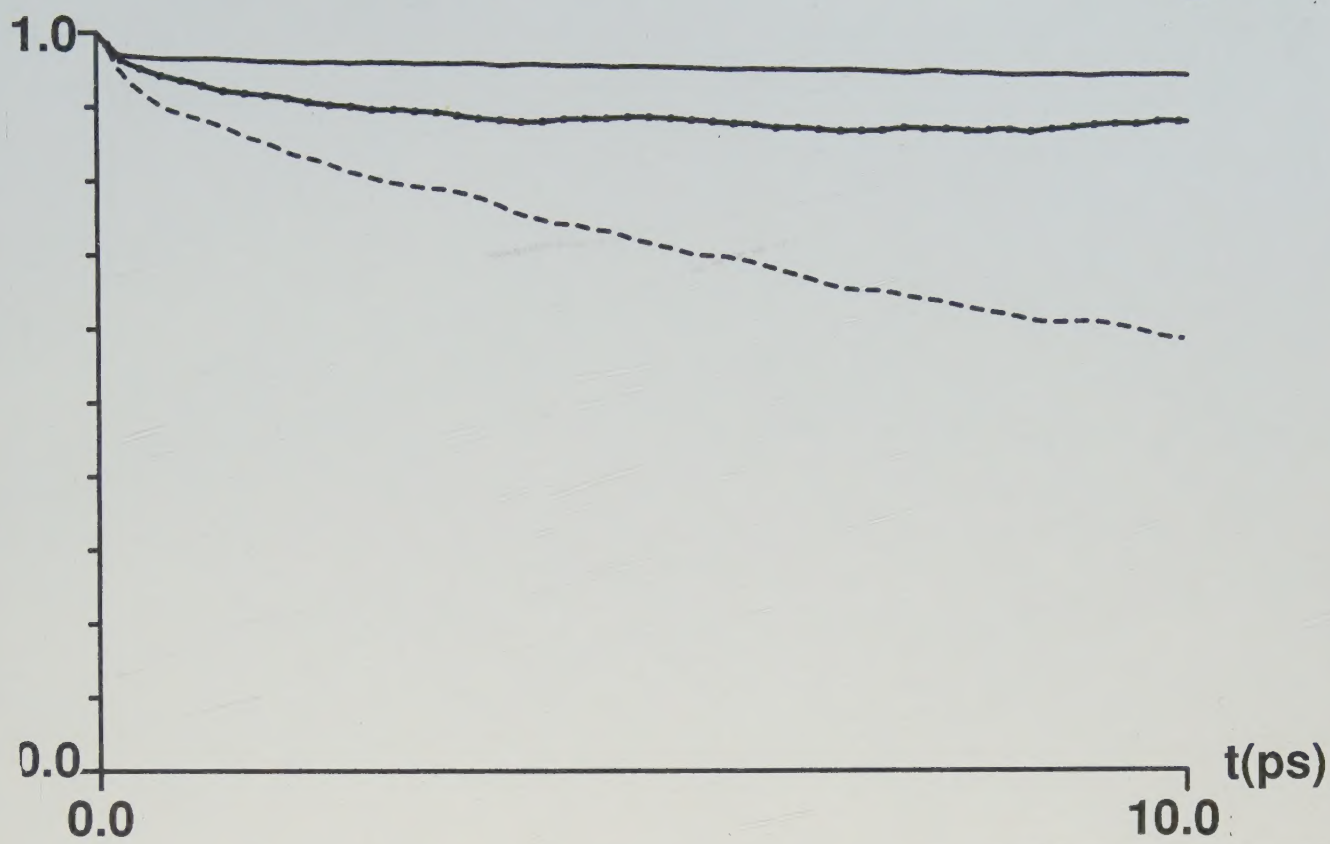


Fig. 9b

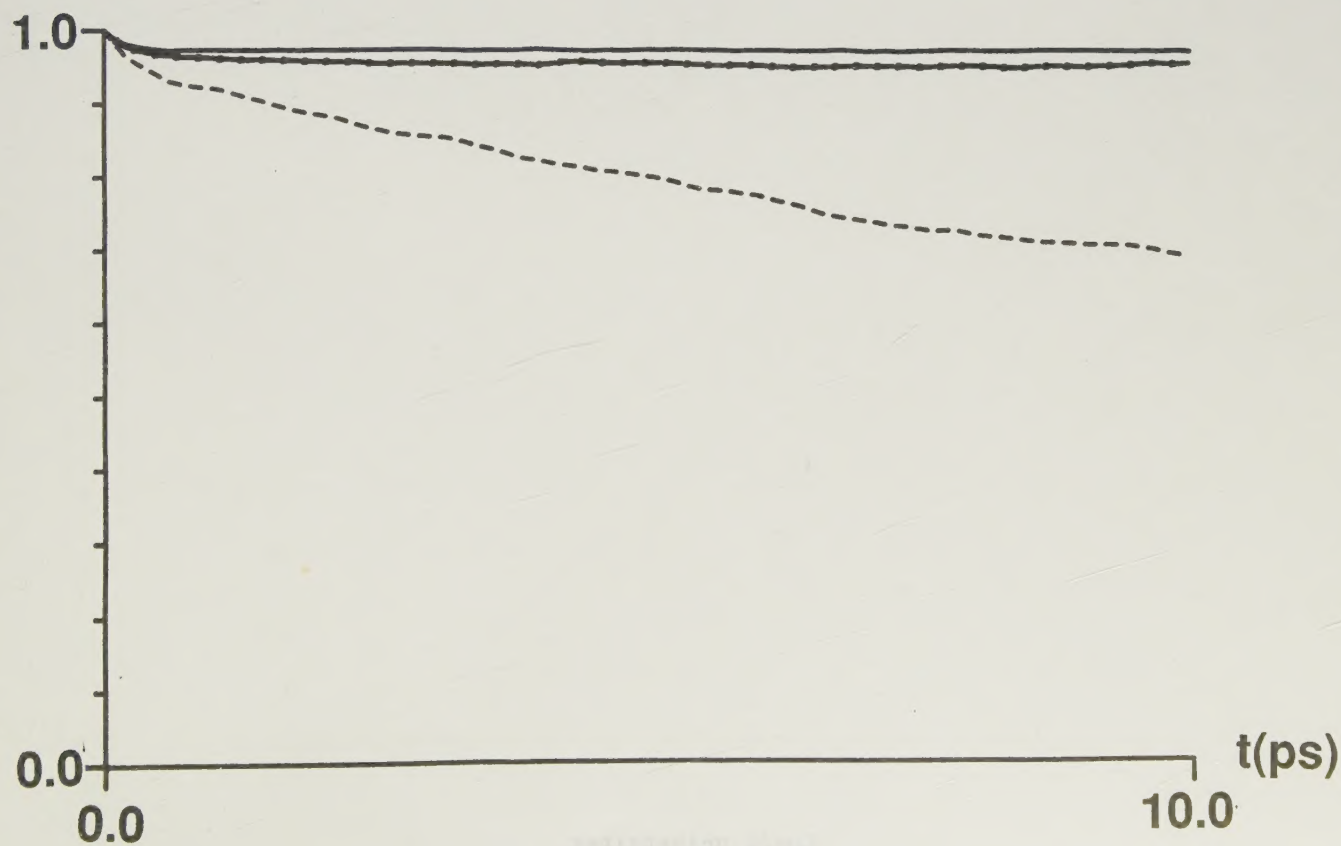
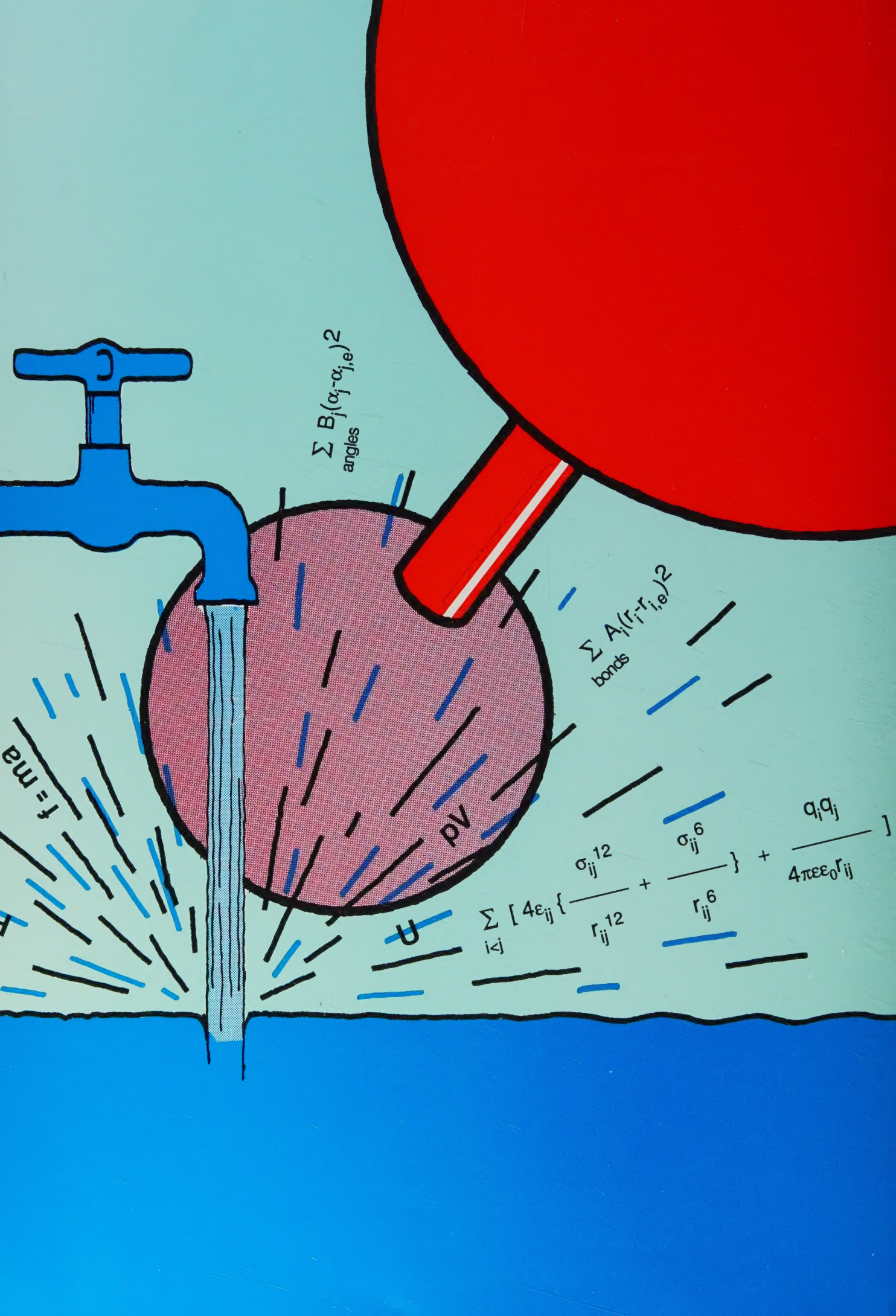


Fig. 8a



Fig. 8b





$$\sum_{\text{angles}} B_j(\alpha_j - \alpha_{j,e})^2$$

$$\sum_{\text{bonds}} A_j(r_j - r_{j,e})^2$$

PV

U

$$\sum_{i < j} [4\epsilon_{ij} \left\{ \frac{\sigma_{ij}^{12}}{r_{ij}^{12}} + \frac{\sigma_{ij}^6}{r_{ij}^6} \right\} + \frac{q_i q_j}{4\pi\epsilon\epsilon_0 r_{ij}}]$$

$$f = ma$$